









```
Db      3  ThrThrPheIysAlaAlaAlaValGlnAlaGluProValITrPmetAspAlaAlaThr 22
Qy      76  GTCCAGAAAGGATCGGCTGATCGAGCAGCGCGCAAGCAGCATGCGCTGATCGCA 135
Db      23  AlaAspIysThrValIThrLeuValAlaAlaValAlaAlaValAlaGlnLeuValAla 42
Qy      136  TTCACAGACATTGGATTCCCGGCGATCCCTTTGGATATAGCGTCGGGCGCCGCTTGG 195
Db      43  PheProGluLeuTrpIleProGlyTrpProGlyPheMetLeuThrHisAsnGlnThrGlu 62
Qy      196  GGATCGGCTTCGTCACAGCGCTATTTCGAGATTCGCTCGTCGCGCAGCAGCAGTGG 255
Db      63  ThrLeuProPheIleIleIleIleIleIleIleIleIleIleIleIleIleIleIle 82
Qy      256  CAGGCGCTGGGAGATCGCGCGCGCGCGCGCATGCGATGCTGCGCGCGCTATAGCGAG 315
Db      83  GluIysIleIleArgCysAlaAlaGlnGlnHisAsnIleAlaLeuSerPheGlyTyrSerGlu 102
Qy      316  CGCGCGGCGCGCAGCTCTATATAGCGCAGCATTCCTCGCGCGCGCATGCGATCTGATC 375
Db      103  ArgAlaGlyArgThrLeuIleIleIleIleIleIleIleIleIleIleIleIleIle 122
Qy      376  CGCGCGCGCGCGCAGCTCTAAGCTTACCATCGGAGCGCAGCGCTGTCGCGAGGAGAC 435
Db      123  IleArgIleArgIleIleIleIleIleIleIleIleIleIleIleIleIleIleIle 142
Qy      436  GGCAGCGCATTCGCGGTCAGCATCCGCCATCGGCGCGCTCGCGCGCTGTTGCTGG 495
Db      143  GlySerAspLeuGlnValAlaGlnThrSerValGlyIleArgValAlaAlaAsnCysAla 162
Qy      496  GAGCAGATCCAGCATTCGTGGAATAGCCATGTAACCGCGCGCAGCAACAGCTTCACGC 555
Db      163  GluAsnLeuGlnSerLeuAsnIleIleIleIleIleIleIleIleIleIleIleIle 182
Qy      556  GCGTCGTGCGCAGCTTCAGCGCTTATCGCGCATGCGCTATGCGCTCGGAGCGGAGTGC 615
Db      183  SerIleIleIleIleIleIleIleIleIleIleIleIleIleIleIleIleIleIle 199
Qy      616  AATACCGCGCAGCATTCAGCGCTGAGGCGCGCTGCTACGTCGTCGCGCTGTCG 675
Db      200  Gly---AlaIleAsnGlnValITyrAlaAlaGlnThrIleIleIleIleIleIleIle 218
Qy      676  GCGACCGCTTCGCGGATGATGATCAAGTA-----TTGGGATGTACCGCCGACAGAG 729
Db      219  GlnValValGlyProThrIleIleAlaAlaPheGlnIleGluAspArgIleIleIleIle 238
Qy      730  ATGTTCTTCAAGCGCGCGCGCTTTGTCATGATTTTCGCGCGCGCAGCGCGCGCTG 789
Db      239  GlnIleIleIleIleIleIleIleIleIleIleIleIleIleIleIleIleIleIle 256
Qy      790  GCCGAGCGCGTCCCGCGAGCGCAAGAGGAGCTGCTGTCGCGCATTCGACGTCGCGATG 849
Db      257  SerIleSerLeuSerProThrIleGlnGlyIleAlaIleIleIleIleIleIleIle 276
Qy      850  ATCGCGTTCGCGCAGCGCGCGCGCATTCGCGCGCGCATTCACGCGCGCGCATGACG 909
Db      277  LeuGlnAlaIleIleIleIleIleIleIleIleIleIleIleIleIleIleIleIle 296
Qy      910  CGGCTGCTGCTGAT-----CGACGTCGCGCGCGCGCATGCTGCTGACGCTGAT----- 957
Db      297  SerValSerIleAsnArgGlnProIleAlaValSerGluVal---IleAspSerAsn 315
Qy      958  -----GCCGATTCGAAACCGCAAAACGAGGACAGAGCGAGC 993
Db      316  GlyAspGluAspProArgAlaAlaCysGluPro-----AspGluGlyAsp 330
```

## RESULT 5

JC4212

nitritilase (EC 3.5.5.1) - Comamonas testosteroni

C/Specties: Comamonas testosteroni

C/Date: 14-Nov-1995 #sequence\_rev: 1996-08-Feb-1996 #text\_change 31-Dec-2004

C/Accession: JC4212, PC4056

R:Levy-Schli, S.; Soubrier, F.; Cuitz-Le Coq, A.M.; Faucher, D.; Crouzet, J.; Petre, D.

```
Gene 161, 15-20, 1995
A>Title: Aliphatic nitritilase from a soil-isolated Comamonas testosteroni sp.: Gene clone
A/Reference number: JC4212; MUID:95369726; PMID:7642130
A/Accession: JC4212
A/Molecule type: DNA
A/Residues: 1-354 <LEF>
A/Cross-references: UNIPROT:Q59329; UNIPARC:UPI000008988B; GB:I32589; NID:G1082008; PIDN
A/Accession: PC4056
A/Molecule type: protein
A/Residues: 154-162/295-302/323-340 <LE2>
A/Cross-references: UNIPARC:UPI0000175E32; UNIPARC:UPI0000175E33; UNIPARC:UPI0000175E34
C/Comment: This enzyme is active on adiponitrile and cyanovaleric acid.
A/Gene: nita
C/Superfamily: nitritilase (carbon-nitrogen hydrolase)
C/Keywords: hydrolase
F/163/Active site: Cys #status predicted

Alignment Scores:
Pred. No.: 5,88e-31 Length: 354
Score: 592.50 Matches: 137
Percent Similarity: 54.0% Conservative: 51
Best Local Similarity: 39.4% Mismatches: 133
Query Match: 31.5% Indels: 27
DB: Gaps: 7

US-09-751-299-1 (1-1041) x JC4212 (1-354)
Qy      31  GCGCGCGTGAAGCGCGCGCGCTGTTCTCGATCTCGACCGCAGTCAGAAAGCATC 90
Db      10  AlaAlaValGlnAlaAlaProValPheMetAsnLeuGlnAlaThrValAspIysThrCys 29
Qy      91  GCGCTGATGAGCAGCGCGCGCAAGCAGCATGTCGCGCTGATGCGCATTCGACGACTTGG 150
Db      30  IyLeuIleIleGlnAlaIleAlaSerMetGlyAlaIysValIleGlyPheProGluAlaPhe 49
Qy      151  ATTCCGCGCTATCCCTTTGGATATGCTGGCG-----GCGCGGCTTGG 195
Db      50  IleProGlyTyrProIleTrpIleTrpIleTrpIleTrpIleTrpIleTrpIleTrpIle 69
Qy      196  GCGATCGGCTGTCGAGGCTATTCGAGATTCGAGATTCGCGCGCGCGCGCGCATGCG 255
Db      70  AlaVal-----LeuPheIyAsnAlaIleGlnIleProSerIyGluVal 84
Qy      256  CAGGCGCTGCGAGATCGCGCGCGCGCGCATGTCGTCGCGCGCTATAGCGAG 315
Db      85  GlnGlnIleSerAspAlaIleIleIleIleIleIleIleIleIleIleIleIleIleIle 104
Qy      316  CGCGCGGCGCGCAGCTCTATATGCGCGCGCATTCGCGCGCGCATTCGCGCATGTCGATC 375
Db      105  IyAspAsnAlaSerLeuIyIleuThrGlnIleuTrpPheAspProAsnIyAsnIleuIle 124
Qy      376  GCGCGCGCGCGCAGGCTCAAGCTTACCCATGCGGAGCGCAGCGCATTCGCGCAGGAGAC 435
Db      125  GlyIyAsnIleArgIyPheIySerProThrIleSerGlnAlaValITrpgIyAspGlyAsp 144
Qy      436  GCGACGATCTGCGCGTGCACATACCGCATTCGCGCGCGCTGCGCGCTGCTGCTGG 495
Db      145  GlySerMetAlaProValPheIySerThrIyIyIyIyIyIyIyIyIyIyIyIyIyIyIy 164
Qy      496  GAGCAGATCCAGCATTCGTGGAATAGCCATGTCAGCGCGCGCAGCAACAGGTCACGTC 555
Db      165  GluHisAlaLeuProIleuAsnIleAlaIleMetGlySerLeuAsnGlnGlnIleIleVal 184
Qy      556  GCGTCGTGCGCAGCTTC-----AGCTCTATTCGCGCGCATGCGC 594
Db      185  AlaSerITrProIleAlaPheValProIyGlyAlaValIleSerIyAsnIyIleSerIyVal 204
Qy      595  TATGCGCTCGCAGCGCAGGTCATACCGCGCGCAGCATTCGCGCGCTGAGGCGCGC 654
Db      205  CysAlaSerThrAsnAlaMetIleGlnIleIleSerGlnPheIyAlaIleSerIyGln 224
Qy      655  TCGTACGTCGTCGCGCTGTCGCGCGCGCGCGCGCGCATGTCAGGATGATGATGTCGAT 714
```

Db 225 VAlTYrAlAlIleMeTSerThrAnLeuValGlYInSPMeTlleApmlelleGlyLys 244  
QY 715 ACGCCGACAGAGATGTTCTCAAGCCGCGCGTGGCATGATTTTCGGGCC 774  
Db 245 ASpIuPhSeRlySAAnPhLeuProLeuGlYSerGlYAnThrAlAlleSeRAsn 264  
QY 775 GAGCGCGCGCCCTGGCGGAGCCGCTCCGAGACCGAAGAGACACTGCTGTGGCCGAT 834  
Db 265 ThrGlYInIleuAlaSer--IleProGlnAplAglInglYIleAlaValAglu 283  
QY 835 ATGACCTCGGACATGATCGCTTGCCAGAGCGGCGGATCGCGGCGGCACTATTCA 894  
Db 284 ILeAspLeuAnGlnIleIleTyrlYlSerTrpLeuApmProAlAglYIleTySer 303  
QY 895 CGGCGGACGTAAAGCGGCTGCTGCTGATGACGT-----CGGCGGCAAGCGGTC 945  
Db 304 ThrProGlnPheLeuSerLeuThrPheApmInserGlnIleValProValYlYleIle 323  
QY 946 -----GTCAAGCTTGATGCGGATCGGATCGAACCAGAAACAGACAGGCGAC--- 993  
Db 324 GlYInglInThrAnIleSphelleSerTyrluApmleuHISgluApmLysMetAspMet 343  
QY 994 -----CGCGCGCGCTGCGCGT 1011  
Db 344 LeuThrIleProProAArgVal 351

## RESULT 6

S77025  
nitrilase (EC 3.5.5.1) - *Synechocystis* sp. (strain PCC 6803)  
N/Alternate names: protein s110784  
C/Species: *Synechocystis* sp.  
A/Variety: PCC 6803  
C/Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 31-Dec-2004  
C/Accession: S77025  
R/Kanehiko, T.; Sato, S.; Korani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;  
O, K.; Okumura, S.; Shimo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda  
DNA Res. 3, 109-136, 1996  
A/Title: Sequence analysis of the genome of the unicellular cyanobacterium *Synechocystis*  
s.  
A/Reference number: S74322; MUID:97061201; PMID:8905231  
A/Accession: S77025  
A/Status: nucleic acid sequence not shown; translation not shown  
A/Molecule type: DNA  
A/Residues: 1-346 <KAN>  
A/Cross-references: UNIPROT:Q05949; UNIPARC:UPI000007133; EMBL:D64005; GB:AB001339; NID  
A/Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996  
C/Genetics:  
A/Start codon: GTG  
C/Superfamily: nitrilase (carbon-nitrogen hydrolase)  
C/Keywords: hydrolase

## Alignment Scores:

Pred. No.:	1.94e-30	Length:	346
Score:	584.50	Matches:	128
Percent Similarity:	59.5%	Conservative:	53
Best Local Similarity:	42.1%	Mismatches:	114
Query Match:	31.1%	Indels:	9
DB:	2	Gaps:	6

US-09-751-299-1 (1-1041) x S77025 (1-346)

QY 25 CGCGGCGCGGCGGCGGCGCGCGCGGCTGTTCTCGATTCGACCGCAAGTCGAGAA 84  
Db 14 ArgAlaIleAlaIleAglInIleSerProValIleuPheSerGlnGlnIleThrMeGluLys 33  
QY 85 GCGATCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 144  
Db 34 ValIleuApmAlaIleAlaIleApmAlaIleValYlYlAgluIleValIlePheProGln 53  
QY 145 ACTTGATTCGCGGCTATCCCTTTGGATATGCGTGCGCGCGCGGCT---TGCGGATG 201  
Db 54 ThrPheValProTyrlYrProTyrlYrPheSerPheValIleuProProValIleuMeGlyLys 73

QY 202 CGTTGCTCCAGGCTATTTCGAGATTGCGTGTGCGGCGGCAAGCAGTGGCAGGCC 261  
Db 74 SerHleuIleuIleuTyrlGlnIleAlaValThrValProIly--LysValThrGlnAla 92  
QY 262 CTGGCGGATGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 321  
Db 93 ILeAlGlnIleAlaIleTyrlThrIleSglYMetValValIleuIleValApmGlnAglu 112  
QY 322 GCGGCGGCGGCTGATATGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 381  
Db 113 GluGlYSerLeuTyrlAnThrGlnIleuIlePheApmAlaApmIleValIleuValLys 132  
QY 382 CGCGGAGAGTCAAGCTTACCATCGGAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 441  
Db 133 ArgArgIleThrProThrTyrlHISgluArgMetValTrpIleGlnIleApmGlyAla 152  
QY 442 CATCTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 501  
Db 153 GlYLeuArgThrValApmThrThrValGlyArgIleuIleAlaIleuAlaCysTrpGlnHIS 172  
QY 502 ATCCAGCATTTGCAATATGCGCATGACGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 561  
Db 173 TyrApmProLeuAlaArgTyrlAlaIleuMeValGlnHISgluIleIleHISgluGln 192  
QY 562 TGGCGGAGCTTCAAGCTTATCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 621  
Db 193 PheProGlySerMetVal-----GlyGlnIlePheAlaApmIleuMetGluVal----- 208  
QY 622 GCGCGAAGCAGATCTAAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 681  
Db 209 ---ThrMetArgHISleIleAlaIleuGlnSerGlyCysPheValIleApmAlaThrGlyTrp 227  
QY 682 GTTTCGCGGAGATGATCAAGTATTTGGTATGATGCGGCGGCGGCGGCGGCGGCGGCGG 741  
Db 228 LeuThrAlaGluIn---LysIleuGlnIleThrThrApmIleuMetHISGlnAlaIleu 246  
QY 742 GCGCGGCGGCGGTTTGGCATGATTTTCGGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 801  
Db 247 SerGlyGlyCysArgThrAlaIleIleSerProGlnIleuIleuIleuIleuIleuIleu 266  
QY 802 CGGAGACGGAAGAGGAGCTGCTGCGGCGGATGACCTCGGATGATGCGGCGGCGGCGG 861  
Db 267 AlaGln---GlyGluIleuAlaIleAlaIleuApmIleuApmIleuAlaIleuArg 285  
QY 862 AAGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 921  
Db 286 LysArgMetMetApmSerValGlyHISTyrlAlaArgProApmIleuIleuIleuIleu 305  
QY 922 GATGACATGTCG 933  
Db 306 AsnApmGlnPro 309

## RESULT 7

JQ1613  
cyanide hydratase (EC 4.2.1.66) - imperfect fungus (*Gloeocercospora sorghi*)  
C/Species: *Gloeocercospora sorghi*  
C/Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 31-Dec-2004  
C/Accession: JQ1613  
R/Wang, P.; VanBreen, H.D.  
Biochem. Biophys. Res. Commun. 187, 1048-1054, 1992  
A/Title: Cloning and properties of a cyanide hydratase gene from the phytopathogenic fun  
A/Reference number: JQ1613; MUID:92412068; PMID:1382413  
A/Molecule type: DNA  
A/Residues: 1-368 <WAN>  
A/Cross-references: UNIPROT:P32964; UNIPARC:UPI0000128C2B; GB:M99044; NID:g168196; PIDN:  
C/Comment: This enzyme converts HCN to formamide.  
C/Genetics:  
A/Gene: Cht  
C/Superfamily: nitrilase (carbon-nitrogen hydrolase)  
C/Keywords: carbon-oxygen lyase; glycoprotein; hydro-lyase  
F/286/Binding site: carbonylhydrate (Asn) (covalent) #status predicted



```

Db      73 -----ArgGlyPheArgPheGlyLeuAlaValGlyValHisAsnGluGluGlyArgAsp 90
Qy      208 ---GTCAGCGCTATTTCGAGAAATTCGCTCGTGGCGGAGCAAGCAGTGGCAGGCCCTG 264
Db      91 GluPheArgAsnTyrrHisAlaSerAlaIleValProGlyProGluValGluArgLeu 110
Qy      265 GCGGATCGGCGCGCGCGCGCGAGCGGAGCGGCGCTGAGCGCTTACGAGCGCGCGCGG 324
Db      111 AlaGluLeuAlaGlyValAsnAsnValHisLeuValHisGlyAlaIleGluValAspGly 130
Qy      325 GCGAGCTCTATATGGCGCGAGCGCATCTCGCGCCGATGGCGCATGTGATCGCGCGCGC 384
Db      131 TyrrHisLeuTyrrCysThrAlaLeuPhePheSerProGluGluGlnPheLeuGlyHis 150
Qy      385 CGCAAGCTCAAGCTTACCTCCATCGCGAGCGCAGCGCTGTTGGCGGAGGAGAGCGGACCAT 444
Db      151 ArgValMetProThrSerLeuGluArgCysIleTrrGlyGlnGlyAspGlySerThr 170
Qy      445 CTGCGCGGTGACGAGTACCGCATGCGCGCGCGCTCGCGCGCTGTTGCTGGAGACATC 504
Db      171 IleProValTyrrAspThrProIleGlyValIleGlyAlaIleCysTrrGluAsnArg 190
Qy      505 CAGCATTTGCTGAATAGATGACCATGTACGCGCGCGAGCAAGCATGCTCAGCTCGCTGG 564
Db      191 MetProLeuTyrrArgThrAlaLeuTyrrAlaValGlyIleGluIleTyrrCysAlaProThr 210
Qy      565 CCGAGCTTCAGCTCTATCGCGGAGTGGCTTATGCGCTCGGAGCGGAGTCAATACCGCC 624
Db      211 AlaAspTyrrSerLeu-----GluTrrGln 218
Qy      625 GCAAGCCAGATCTAC---GCGGTGAGGCGCGGCTCTACGTCGTCGCGTGC----- 672
Db      219 AlaserrMetIleHisIleAlaValGluGlyGlyCysPheValLeuSerAlaHisGlnPhe 238
Qy      673 TGCGGACGCGTTTCG---CCGAG-----ATGATCAAGTATTGATGATACG 717
Db      239 CysValArgArgGluPheProGluHisProAspTyrrLeuPheAsnAspIleValAspThr 258
Qy      718 CCCGACAGAGATGTTCTTCMAAGCGCGCGCGGTTTTCGATGATTTTCGAGCCGAC 777
Db      259 LysGluHisAspProThrValSerGlyGlyGly-----SerValIleIleSerProLeu 276
Qy      778 GCGCGCGCGCTGGCGGAGCGGCTCCCGAGACCGAAGAGGACCTGCTGCGCCGATAC 837
Db      277 GlyValValLeuAlaGlyProAsnTyrrGluSer---GluGlyLeuValIleThrAlaAspLeu 295
Qy      838 GACCTCGCATGATCGGCTTGGCGGAGCGCGCGCATCGCGGCGGCGCATTTTCACG 897
Db      296 AspLeuValAspIleAlaValGlyLeuTyrrPheAspValValGlyHisTyrrSerLeu 315
Qy      898 CCCGACGTAACGCGGCTGCTGATGATGACGTCGCGCGCAACCGCTGTCACGCTTGAT 957
Db      316 ProAspIlePheAsnLeuThrValAsnGluHisProLysProValThrPheMetThr 335
Qy      958 GCGCGATTCCAGACCGCAACGAG 981
Db      336 LysValGluValAsnGluAspGlu 343

RESULT 9
T49147
nitrlaase (EC 3.5.5.1) 1 [imported] - Arabidopsis thaliana
N/Alternate names: protein T10D17.100
C/Species: Arabidopsis thaliana (mouse-ear cress)
C/Date: 02-Jun-2000 #sequence revision 02-Jun-2000 #text_change 31-Dec-2004
C/Accession: T49147; T52260; T52263
R/D/Author: M.; Vezzi, A.; Modesto, D.; Bigazzi, M.; Valle, G.; Mewes, H.W.; Rudd, S.; I
submitted to the Protein Sequence Database, April 2000
A/Reference number: Z25017
A/Accession: T49147
A/Status: Preliminary
A/Molecule type: DNA
A/Residues: 1-346 <DNA>
A/Cross-references: UNIPROT:P32961, UNIPARC:UPI0000001AE2, EMBL:AL353865, GSPDB:GN00061,

```

```

A/Experimental source: cultivar Columbia; BAC clone T10D17
R/Hilbrand, H.; Bartling, D.; Weiler, E.W.
Plant Mol. Biol. 36, 89-95, 1998
A/Title: Structural analysis of the nit2/nit1/nit3 gene cluster encoding nitrlaases, enz
A/Reference number: Z26007; M0ID:98145459; PMID:9484465
A/Accession: T52260
A/Status: Preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-346 <HL>
A/Cross-references: UNIPARC:UPI0000001AE2; EMBL:Y07648; PIDN:CAA6935.2
A/Experimental source: cultivar Columbia
R/Bartel, B.; Pink, G.R.
Proc. Natl. Acad. Sci. U.S.A. 91, 6649-53, 1994
A/Title: Differential regulation of an auxin-producing nitrlaase gene family in Arabidops
A/Reference number: Z24515; M0ID:8022831; PMID:8022831
A/Accession: T52263
A/Status: Preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-346 <BAR>
A/Cross-references: UNIPARC:UPI0000001AE2; EMBL:U08845; PIDN:AAB05221.1
A/Experimental source: cultivar Columbia
C/Genetic:
A/Gene: ATSP:T10D17.100; nit1; NIT1
A/Map position: 3
A/Intons: 44/1; 104/1; 202/1; 296/1
C/Superfamily: nitrlaase (carbon-nitrogen hydrolase)
C/Keywords: hydrolase

Alignment Scores:
Pred. No.:      8 44e-23      Length:      346
Score:          467.00      Matches:      121
Percent Similarity: 48.2%      Conservative: 50
Best Local Similarity: 34.1%      Mismatches: 122
Query Match:    24.8%      Indels:      62
DB:             2          Gaps:      8

US-09-751-299-1 (1-1041) x T49147 (1-346)

Qy      10  CCATAGCAGATGATGCGGCGCGGCTGACGCGCGCGGTTCTTCATTCGAC 69
Db      21  ProSerThrThrValArgValThrIleValGlnSerSerThrValTyrrAsnAspThrPro 40
Qy      70  CGAGAGTGAGAGAGATGCGGCTGATGACGACGCGCGCGCAAGACGATGCGCGCTG 129
Db      41  AlatrThrIleAspLysAlaGluTyrrIleValGluAlaIleSerIysGlyAlaGluLeu 60
Qy      130  ATGCAATCCCAAGACTTGATCCCGGCTATCCCTTTTGATATGCTGCGCGCGCG 189
Db      61  ValLeuPheProGluGlyPheIleGlyTyrrPro----- 72
Qy      190  GCTTGGGCGATGCGCTTC----- 267
Db      73  ---ArgGlyPheArgPheGlyLeuAlaValGlyValHisAsnGluGluGlyArgAspGlu 91
Qy      208  GTCAGCGCTATTTCGAGAAATTCGCTCGTGGCGGAGCAAGCAGTGGCAGGCCCTGCGG 267
Db      92  PheArgLysTyrrHisAlaSerAlaIleHisValProGlyProGluValAlaArgLeuAla 111
Qy      268  GATCGGCGCGCGCGCGCGCATGTCGTCGCGCGCTGATGCGGAGCGCGCGCGCGC 327
Db      112  AspValAlaArgLysAsnHisValTyrrLeuValMetGlyAlaIleGluValAspGlyTyrr 131
Qy      328  AGCCTCTATATGGCGAGCGGATCTTCGCGCCGATGCGCATGTCGCGCGCGCGC 387
Db      132  ThrLeuTyrrCysThrValLeuPhePheSerProGluGlyGlnPheLeuGlyHisAsnArg 151
Qy      388  AACCTCAAGCTTACCATGCGGAGCGCGCTTGGCGGAGGAGAGCGGACCATCTTC 447
Db      152  LysLeuMetProThrSerLeuGluArgCysIleTrrGlyGlnGlyAspGlySerThrIle 171
Qy      448  GCGGTGACGATACCGCATCGGCGCGCTGCGCGCGCTGTTGCTGGAGACATTCAG 507
Db      172  ProValTyrrAspThrProIleGlyLysLeuGlyAlaAlaIleCysTrrGluAsnAsnMet 191

```



Db 275 G1yProAsnPhenGluSer---GluGlyLeuIleThraIaAspLeuAspLeuGlyAspVal 293  
Qy 853 GCCTGGCCCAAGGCGGCGCGGATCCGCGCGCCATTCATTCACGCGCCGACGTAACCGG 912  
Db 294 AlaArgAlaLeuValLeuTyrPheAspSerValGlyHisTyrSerArgProAspValLeuHis 313  
Qy 913 CTGCTGCTGGATGCACCTCCGCGCGCGATGCTGCACGCTGATGATCGGATTCGACCG 972  
Db 314 LeuThrValIleGlnHisPro--LysLeuProValThrPheIleSerLysValGlyLys 332  
Qy 973 CAAACGAGGAC 984  
Db 333 AlaGluAspAsp 336  
RESULT 11  
T52259  
nitriIase (EC 3.5.5.1) 2 (imported) - Arabidopsis thaliana  
C/Species: Arabidopsis thaliana (mouse-ear cress)  
C/Date: 20-Oct-2000 #sequence\_revision 20-Oct-2000 #text\_change 31-Dec-2004  
C/Accession: T52259  
R/Hillebrand, H.; Bartling, D.; Weiler, E.W.  
Plant Mol. Biol. 36, 89-99, 1998  
A/Title: Structural analysis of the nit2/nit1/nit3 gene cluster encoding nitriIases, enz  
A/Reference number: 226007; PMID:98145459; PMID:9484465  
A/Accession: T52259  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: DNA  
A/Residues: 1-339 <HL>  
A/Cross-references: UNIPROT:O04907; UNIPARC:UP100000A5899; EMBL:Y07648; P1DN:CAA68934.3  
A/Experimental source: cultivar Columbia  
A/Genes: nit2  
A/Intons: 37/1; 97/1; 195/1; 289/1  
C/Superfamily: nitriIase (carbon-nitrogen hydrolase)  
C/Keywords: hydrolase  
Alignment Scores:  
Pred. No.: 1.53e-22 Length: 339  
Score: 463.00 Matches: 119  
Percent Similarity: 50.3% Conservative: 54  
Best Local Similarity: 34.6% Mismatches: 127  
Query Match: 24.6% Indels: 44  
DB: 2 Gaps: 8  
US-09-751-299-1 (1-1041) x T52259 (1-339)  
Qy 16 ACGAAGTATCGCGCGCGCGCGCGCGCGGTGTTCTCGATTCGACCGACA 75  
Db 16 ThrlleValaIrgAlaThrlleValaGlnAlaSerThraValTyrAsnAspThrProAlaThr 35  
Qy 76 GTGAGAAACGATCGCGCTGATGCAGAGCGCGCGCAAGACGATGCGCTGATGCA 135  
Db 36 LeuGluLysAlaAsnLysPheIleValaGlnAlaSerLysGlySerGluLeuVal 55  
Qy 136 TTCCCAAGACTGATTCGCGGTATCC-----TTTGATATGCTGCGCGCGCG 169  
Db 56 PheProGluAlaPheIleGlyGlyTyrProArgGlyPheArgPheGlyLeuGlyValGly 75  
Qy 190 GCTTGG-----GGCATGCGCTTGCATGCGCGCTATTCGAGATTCGCTGTCGC 240  
Db 76 ValHisGlnGluGluGlyArgAspGluPheArgLysTyrHisAlaSerAlaIleLeuVal 95  
Qy 241 GGCAGACAGATGGAGCGCTGGAGATGCGCGCGCGCGCGCGCGCATGATGTCGTG 300  
Db 96 ProGluProGluValGluLysLeuAlaGluLeuAlaGlyLysAsnAsnValTyrLeuVal 115  
Qy 301 GCCGCTATACGCGCGCGCGCGCGCGCGCGCTATGATGCGCGCGCGCGCGCGCG 360  
Db 116 MetGlyAlaIleGlyLysAspGlyTyrThrLeuTyrCysThrAlaLeuPheSerPro 135  
Qy 361 GATGCGCATGATGCG 420  
Db 136 GlnGlyGlnPheLeuGlyLysHisArgLysLeuMetProThrSerLeuGluValGlyCysIle 155

Qy 421 TTGCGAGGAGACG 480  
Db 156 TrpGlyGlnLeuValAspLysLeuThrIleProValTyrAspThrProIleGlyLysLeuGly 175  
Qy 481 GCGCTGTTGCTGGAGACATCCAGCATTCGTGCAATACCGCATGTAACCGCGCGAC 540  
Db 176 AlaAlaIleCysTrpGluAsnArgMetProLeuTyrArgThrAlaLeuTyrAlaGly 195  
Qy 541 GAACAGTCCAGCTCGCG-----TCGTGCCGAGCTTCAGCTTC 579  
Db 196 IleGluLeuTyrCysAlaProThrAlaAspGlySerLysGluTrpGlnSerMetLeu 215  
Qy 580 TATCGGCGCATGCGCTATGCGCTCGAGCGGAGGTCAATACCGCGCGCGCGCGCG 639  
Db 216 His-----Ile 217  
Qy 640 GCGCTCAGGCGCGCGCTGATCGTGTGCGGTGCGCGCGCGCGCGCGCGCGCGCG 699  
Db 218 AlaIleGluGlyGlyCysPheValIleSerAlaCysGlnPheCys-----LeuArg 234  
Qy 700 AACGTATGGTGAACGCC-----GACAGAGATG 732  
Db 235 LysAspPheProAspHisProAspTyrLeuPheThrAspTrpTyrAspAspLysGluPro 254  
Qy 733 TTCTCAGCG 792  
Db 255 AspSerIleValSerGlnGlySerValIleIleSerProLeuGlyGlnValLeuAla 274  
Qy 793 GACCGCTCCGCGAGACCGAAGAGGAGTGTGTCGCGCATTCGACCTCGCGCATGATC 852  
Db 275 G1yProAsnPhenGluSer---GluGlyLeuIleThraIaAspLeuAspLeuGlyAspVal 293  
Qy 853 GCCTGGCCCAAGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 912  
Db 294 AlaArgAlaLeuValLeuTyrPheAsp\*\*ValGlyHisTyrSerArgProAspValLeuHis 313  
Qy 913 CTGCTGCTGGATGCACCTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 972  
Db 314 LeuThrValaIleGlnHisPro--LysLysProValThrPheIleSerLysValGlyLys 332  
Qy 973 CAAACGAGGAC 984  
Db 333 AlaGluAspAsp 336  
RESULT 12  
S22398  
nitriIase (EC 3.5.5.1) - Arabidopsis thaliana  
C/Species: Arabidopsis thaliana (mouse-ear cress)  
C/Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 31-Dec-2004  
C/Accession: S22398  
R/Bartling, D.; Seedorf, M.; Mithoefler, A.; Weiler, E.W.  
Eur. J. Biochem. 205, 417-424, 1992  
A/Title: Cloning and expression of an Arabidopsis nitriIase which can convert indole-3-ac  
A/Reference number: S22398; PMID:92209532; PMID:1555601  
A/Accession: S22398  
A/Molecule type: mRNA  
A/Residues: 1-346 <BAR>  
A/Cross-references: UNIPROT:P32961; UNIPARC:UP10000130513; EMBL:X63445; NID:G16399; P1DN:  
C/Superfamily: nitriIase (carbon-nitrogen hydrolase)  
C/Keywords: hydrolase  
Alignment Scores:  
Pred. No.: 2.07e-22 Length: 346  
Score: 461.00 Matches: 120  
Percent Similarity: 48.2% Conservative: 51  
Best Local Similarity: 33.8% Mismatches: 122  
Query Match: 24.5% Indels: 62  
DB: 2 Gaps: 8  
US-09-751-299-1 (1-1041) x S22398 (1-346)  
Qy 10 CCATACAGAAATATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 69

RESULT 13

152262

nitrilase (EC:5.5.5.1) 2 [imported] - Arabidopsis thaliana  
C.Species: Arabidopsis thaliana (mouse-ear cress)  
C.Date: 20-Oct-2000 #sequence\_revision 20-Oct-2000 #text\_change 31-Dec-2004  
C.Accession: T52262  
R.Barrett, B.; Fink, G.R.  
Proc. Natl. Acad. Sci. U.S.A. 91, 6649-53, 1994  
A.Title: Differential regulation of an auxin-producing nitrilase gene family in Arabidopsis  
A.Reference number: Z24515; MUID:8022831; PMID:8022831  
A.Accession: T52262  
A.Status: Preliminary; translated from GB/EMBL/DDJJ  
A.Molecule type: DNA  
A.Residues: 1-339 <BAR>  
A.Cross-references: UNIPROT:P32962; UNIPARC:UPI000016DA4C; EMBL:U38845; P1DN:AA05220.1  
A.Experimental source: cultivar Columbia  
C.Genetics:  
A.Gene: NIT2  
A.Map position: 3  
A.Introns: 37/2; 97/1; 195/1; 289/1  
C.Superfamily: nitrilase (carbon-nitrogen hydrolase)  
C.Keywords: hydrolase

Alignment Scores:	
Pred. No.:	4.38e-22
Score:	456.00
Percent Similarity:	50.0%
Best Local Similarity:	34.3%
Query Match:	24.2%
DB:	2
US-09-751-299-1 (1-1041) x T52262 (1-339)	
Length:	339
Matches:	118
Conservative:	54
Mismatches:	128
Indels:	44
Gaps:	8

OY		16	ACGAAGTTCGGGCGGGCGGTGCAGGCGGGCGGGATCTCCGATCTCGAACCGACA	75
Dd		16	ThrlleValdaIgaIaThrllleValalGlmlaserlhnValTyrrAemphrPrroalaThr	35
OY		76	GTCGAGAAGGCGATGGCCCTGATCGACGAGCGGCCAACGACGACTGGCCTGATCGCA	135
Dd		36	LenglylvlaAlaaemLysPheIlleValalGlmlaalhmLlysGlyserglueuValval	55
OY		136	TTCSCAGAGCTTGATGCCGCGCATGCC-----TTTGGATATGGCTGGGCGGCGG	189
Dd		56	PheProgluaIaphellegLyglYtyrProargglYpheArgPheoIyLeuglyValagly	75
OY		190	GCTTGG-----GGCATGCGCTTGCTCCAGCCCTATTGGAGAATTCCGCTGTCGCC	240
Dd		76	ValHlsamengluglYlYtArgspdlurPheArgglyTYrhAlaserAlallelysvaI	95
OY		241	GGCAGCAAAGTGGGAGGCCCTCGGAGTGCAGCGCGCGCGCACGGCATGCAATGCTGTA	300
Dd		96	ProglYprogluValgluLYslvleuhalagluenualaglYlyvaAmhavalTYzlvauaI	115
OY		301	GCGCGGTAATGAGGCGCGCGCGCGGCGAGCCCTTATGGCGGAGGCGATCTTCGGCCC	360
Dd		116	MercilYalallleglYlvAsprlYTYrThreuuYCYsthAlaleudhePheSerPro	135
OY		361	GATGGCGATGTGATCGCGCGCGCGCGCAAGCTCAAGCCTTACCATCGGAGGCGACCGTA	420
Dd		136	GlmlYglInPheleuaglYlvnHlsArglyvleuMetProthrSerleuclunArgvyle	155
OY		421	TTTCGGCAGGAGAGCGGAGCCATCTCGCGGCGGAGCAAGTACCGGCATCGGGCGCTCGGC	480
Dd		156	TtpolYgenlglyAsprgIyserThrIleProvalITyzabhrmProIleelylvseuagly	175
OY		481	GGCGTCTGTGCTGGGAGACATCGACATCGCATTTGCAATAACGCGATGACCGCGCGAC	540
Dd		176	AlaAlaIleCystrpgluAsnaHgmecProleuTYrAgThralaleuYrzAlavsgly	195
OY		541	GAACAGGTCCAAGTCGCGG-----TCGTGCCAGGCTTCAGCCCTC	579
Dd		196	IleglueuYrysvaIaprotHrAlaasprgIyserlysvglutrpclnsersetmetleu	215
OY		580	TATCGCGGCAATGGCCTTATGCGCTCGAGCCGAGAGTCAATACCGCGCAAGCCAGATCTAC	639



216 His-----11e 217  
QY 640 GCGGTGAGGCGGCTGCTACGTGCGGTGTCGCGACCGCTTTCGCGAGATGATC 699  
Db 218 AlaIleGluGlyGlyCysPheValIleuSerAlaCysGlnPheCys-----LeuArg 234  
QY 700 AAGGATTTGGTGGATACGCCC-----GACAGAGGATG 732  
Db 235 LysAspRheProAspRhiProAspTyrIleuPheThrAspTyrTyrAspAspLysGluPro 254  
QY 733 TTCCTCAAGGCGCGGCGGCTTTTGCATGATTTTCGAGCCGACGCGCGCGCTGCGC 792  
Db 255 AspSerIleValSerGlnGlyGlySerValIleIleSerProLeuGlyGlnValLeuAla 274  
QY 793 GAGCGGCTCCGAGACCGAAGAGGAGTCTGCTGTCGCGGATATGACCTGGCATGATC 852  
Db 275 GlyProAspRheGluSer---GlnGlyLeuIleThrAlaAspLeuAspLeuGlyAspVal 293  
QY 853 GCGTTGCGCAAGCGCGCGGCGGATCCGCGGCGCATTTACACGCGCGCGGATACGCGG 912  
Db 294 AlaArgAlaLysLeuTyrPheAspSerValGlyHisTyrSerArgProAspValLeuHis 313  
QY 913 CTGCTGTGATTCGACGTCCGCGCCGACGCGTGTGACGCTTGATGCGACATTCGACCG 972  
Db 314 LeuThrValaGlnIlePro---LysLysProValThrPheIleSerLysValGluLys 332  
QY 973 CAACACGAGAC 984  
Db 333 AlaGluAspAsp 336  
RESULT 14  
T52266  
nitrlase-like protein [imported] - rice  
C/Species: Oryza sativa (rice)  
C/Date: 20-Oct-2000 #sequence\_revision 20-Oct-2000 #text\_change 31-Dec-2004  
C/Accession: T52266  
R/Chiba, R.; Dohmoto, M.; Yamaguchi, K.  
submitted to the EMBL Data Library, May 1999  
A/Description: Oryza sativa, a gene for nitrlase-like protein.  
A/Reference number: Z26008  
A/Accession: T52266  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: mRNA  
A/Residues: 1-362 <CHT>  
A/Cross-references: UNIPROT:Q9SXK6; UNIPARC:UPI000009C92A; EMBL:AB027054; PIDN:BAA77679.  
A/Experimental source: cultivar Nipponbare  
C/Genetics:  
A/Gene: ONIT4  
C/Superfamily: nitrlase (carbon-nitrogen hydrolase)  
Alignment Scores:  
Pred. No.: 9.61e-20 Length: 362  
Score: 420.00 Matches: 117  
Percent Similarity: 50.7% Conservative: 57  
Best Local Similarity: 34.1% Mismatches: 137  
Query Match: 22.3% Indels: 32  
DB: 2 Gaps: 10  
US-09-751-299-1 (1-1041) x T52266 (1-362)  
QY 25 CGCGCGCGGCGGTGTCAGACCGCGCGGTTCCTGATTCGACCGACGACGTCAGAAA 84  
Db 32 ArgAlaThrValValGlnAlaSerThrValPheTyrAspThrProAlaThrLeuAspLys 51  
QY 85 GCGATCGGCTGATCGACGCGCGCGCAAGACGACGTCGTCGATGATTCGACG 144  
Db 52 AlaGluValGluIleGluAlaAlaGlyTyrGlySerGlnLeuValValPheProGlu 71  
QY 145 ACTTGATTCGCGGCTATCCCTTTGG-----ATATGCGTG 180  
Db 72 AlaPheValGlyTyrProArgLysSerThrPheGlyPheGlyAlaAsnIleSerIle 91  
QY 181 GCGCGCGCGGCT---TGCGGACATGCGCTTCGTCACGCGCTATTCGAGAAATTCGCTCGTG 237

Db 92 GlyAsnProLysAspLysGlyLysGluGluPheArgLysTyrHisAlaAlaIleGlu 111  
QY 238 CGCGGACGACGACGTCGACAGGCTTGCGGATGCGCGCGCGCGACGATGATC 297  
Db 112 ValProGlyProGluValThrArgLeuAlaIleMetAlaGlyLysTyrLysValPheLeu 131  
QY 298 GTGCGCGCTATACGACCGCGCGCGGCGGACGCTCTATAGGCGCGATTCGCGC 357  
Db 132 ValMetGlyValIleGluArgGluGlyTyrThrLeuTyrCysSerValLeuPhePheAsp 151  
QY 358 CCCGATGCGCATGATGCGCGCGCGCGCGACGTCGACGCTTACGCGACGCGAC 417  
Db 152 ProLeuGlyArgTyrLeuGlyLysHisArgLysLeuMetProThrAlaLeuGluArgIle 171  
QY 418 GTTGTGCGGAGGACGACGCGACGATTCGCGGTCGACGATACGCGCGATCGCGCTC 477  
Db 172 IleTyrGlyPheGlyAspGlySerThrIleProValTyrAspThrProLeuGlyLysIle 191  
QY 478 GCGCGCTCTGCTGTCGCGGACGACATCGACGATTCGGAATCGCATGTACGCGCC 537  
Db 192 GlyAlaLeuIleCysTyrGluAsnLysMetProLeuLeuArgThrAlaLeuTyrGlyLys 211  
QY 538 GACGACAGGTCGACGTGCGCGTGGCGCGGATTCGACGCTGATCGCGGATGCGCTAT 597  
Db 212 GlyIleGluIleTyrCysAla-----ProThr----- 220  
QY 598 GCGCTCGACCGGAGTCATACCGCGCGCAAGCATACGCGGTGACGCGCGCTGC 657  
Db 221 AlaAspSerArgGlnValTyrGlnAlaSerMetThrHisIleAlaLeuGluGlyCys 240  
QY 658 TACGTGCTGCGGTG-----TGCGGACGCTTTCG-----CGGAGATG 696  
Db 241 PheValIleuSerAlaGlnPheCysArgArgLysAspTyrProProProGluTyr 260  
QY 697 ATCAAGGTA---TTGATGATAGCGCGCGACGAGATGTTCTCAAGCGCGCGCGCT 753  
Db 261 ValPheThrGlyLeuGlyGluGluProSerProAspThrValValCysProGlyGly--- 279  
QY 754 TTTCGATGATTTTCGCGCGCGACGCGCGCGCTGCGCGGCTCCGAGACGGA 813  
Db 280 ---SerValIleIleSerProSerGlyGluValLeuAlaGlyProAsnTyrGlu---Gly 297  
QY 814 GAGGACGCTGCTGCGCGCATATGACCTTGCGCATGATTCGCGTTCGCGGCGCGCGC 873  
Db 298 GluAlaLeuIleThrAlaAspLeuAspLeuGlyGluIleValArgAlaLysPheAspPhe 317  
QY 874 GATCGCGCGGCGCATTTACACGCGCGCGGATGCGCGCTGCTGATTCGACGTCGCG 933  
Db 318 AspValValGlyHisTyrAlaArgProGluValIleuSerLeuValIleAspArgIlePro 337  
QY 934 GCCCAAGCGCTGTCAGCTTGATGCGCGCATTCGACCGCAACGCGCAAGACGCGGAC 993  
Db 338 ---HisLeuProValSerPheThrSerAlaGluLysThrThrAlaIleLysSerAsp 356  
QY 994 GCGCGCGCG 1002  
Db 357 SerThrAla 359  
RESULT 15  
T03739  
nitrlase (EC 3.5.5.1) 4B - common tobacco  
C/Species: Nicotiana tabacum (common tobacco)  
C/Date: 24-Mar-1999 #sequence\_revision 24-Mar-1999 #text\_change 31-Dec-2004  
C/Accession: T03739  
R/Tanoda, H.  
submitted to the EMBL Data Library, January 1996  
A/Reference number: Z15044  
A/Accession: T03739  
A/Status: preliminary; translated from GB/EMBL/DBJ  
A/Molecule type: mRNA  
A/Residues: 1-348 <TSU>  
A/Cross-references: UNIPROT:Q42966; UNIPARC:UPI000009F332; EMBL:D83078



GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: April 27, 2006, 01:07:06 ; Search time 51.2142 Seconds  
(without alignments)  
2868.171 Million cell updates/sec

Title: US-09-751-299-1  
Perfect score: 1881  
Sequence: 1 atcgcgcagccatgacgacaa.....ggcgcgcgcgcgcgcagtag 1041

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 216643 seqs, 705528306 residues  
Total number of hits satisfying chosen parameters: 4332886

Minimum DB seq length: 0  
Maximum DB seq length: 200000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-Q=/abses/ABSSWEB.spool/US09751299/runatc\_26042006\_090034\_18631/app\_query.fasta\_1  
-DB=UniProt -QFMT=fastan -SUFFIX=rup -MINMATCH=0.1 -LOOPEXT=0 -LIST=45  
-UNITS-bits -START=1 -END=-1 -MATRIX=bl0sum62 -TRANS=human40.cdi -LIST=45  
-DOCCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pro -NORW=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000 -HOST=abss02p  
-USER=US09751299.@CCN\_1\_1\_580.@runatc\_26042006\_090034\_18631 -NCPU=6 -ICPU=3  
-NO MMAP -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG -DEV TIMEOUT=120  
-WARN TIMEOUT=30 -THREDS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : UniProt 05.80.\*  
1: uniProt\_sprot.\*  
2: uniProt\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1806	96.0	346	2	Q6RMK5_9222Z
2	1464	77.8	341	2	Q6RMW4_9222Z
3	1209	64.3	348	2	Q6RMS2_9222Z
4	1165	61.9	353	2	Q6RMF9_9222Z
5	1155	61.4	353	2	Q6RMF0_9222Z
6	1132.5	60.1	358	2	Q6RMW0_9222Z
7	1130	60.2	354	2	Q6RMW10_9222Z
8	1127	59.9	331	2	Q706C8_PSBPU
9	1125	59.8	345	2	Q6RMS5_9222Z
10	1099.5	58.5	334	2	Q6RME3_9222Z
11	1076.5	57.2	337	2	Q6RMW9_9222Z
12	1073.5	57.1	338	2	Q6RMW6_9222Z
13	1070	56.9	336	2	Q6RMW6_9222Z
14	1068.5	56.8	335	2	Q6RMW7_9222Z
15	1068.5	56.8	336	2	Q5YUM5_NOCFA
16	1066.5	56.7	337	2	Q6RMW8_9222Z

17	1064.5	56.6	338	2	Q6RMW7_9222Z	Q6RMW7 uncultured
18	1064.5	56.6	350	2	Q5EG61_PSEFL	Q5EG61 pseudomonas
19	1057.5	56.2	352	2	Q6RMH0_9222Z	Q6RMH0 uncultured
20	1052.5	56.0	338	2	Q6RME4_9222Z	Q6RME4 uncultured
21	1044.5	55.5	337	2	Q6RMS3_9222Z	Q6RMS3 uncultured
22	1043.5	55.5	338	2	Q6RME2_9222Z	Q6RME2 uncultured
23	1041	55.3	336	2	Q6RMW5_9222Z	Q6RMW5 uncultured
24	1038.5	55.2	338	2	Q6RMW5_9222Z	Q6RMW5 uncultured
25	1038.5	55.2	338	2	Q6RMW6_9222Z	Q6RMW6 uncultured
26	1028.5	54.7	337	2	Q6RMW7_9222Z	Q6RMW7 uncultured
27	1026.5	54.6	338	2	Q6RMW1_9222Z	Q6RMW1 uncultured
28	1025.5	54.5	336	2	Q50UJ1_PSEBY	Q50UJ1 pseudomonas
29	1023	54.4	337	2	Q6RMW3_9222Z	Q6RMW3 uncultured
30	1023	54.4	337	2	Q6RMW7_9222Z	Q6RMW7 uncultured
31	1020.5	54.3	337	2	Q6RMW9_9222Z	Q6RMW9 uncultured
32	1012	53.8	347	2	Q8BB32_PSSBM	Q8BB32 pseudomonas
33	1008	53.6	349	2	Q6RMS0_9222Z	Q6RMS0 uncultured
34	1007	53.5	337	2	Q6RME0_9222Z	Q6RME0 uncultured
35	1004.5	53.4	337	2	Q6RMW3_9222Z	Q6RMW3 uncultured
36	986	52.4	345	2	Q6RMW2_9222Z	Q6RMW2 uncultured
37	974.5	51.8	337	2	Q6RMW9_9222Z	Q6RMW9 uncultured
38	972	51.7	334	2	Q6RMW9_9222Z	Q6RMW9 uncultured
39	964	51.2	334	2	Q89GE3_BRAJA	Q89GE3 bradyrhizob
40	949.5	50.5	346	2	Q6RMW1_9222Z	Q6RMW1 uncultured
41	949	50.5	333	2	Q6RMW0_9222Z	Q6RMW0 uncultured
42	945	50.2	332	2	Q6RMW0_9222Z	Q6RMW0 uncultured
43	869	46.2	356	1	NRLA_AICFA	P20960 alcaligenes
44	829.5	44.1	365	1	NRL2_RHORA	003217 rhodococcus
45	825.5	43.9	383	1	NRL1_RHORA	002068 rhodococcus

## ALIGNMENTS

RESULT 1  
Q6RMK5\_9222Z  
ID Q6RMK5\_9222Z PRELIMINARY; PRT; 346 AA.  
AC Q6RMK5;  
DT 05-JUL-2004 (TEMBLrel. 27, Created)  
DT 05-JUL-2004 (TEMBLrel. 27, Last sequence update)  
DE Nitrilase (EC 3.5.5.7).  
GN ORFNames=BD5086;  
OS uncultured organism.  
OC unclassified; environmental samples.  
OX NCBI\_TaxID=155900;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RX PubMed=15066841; DOI=10.1128/AEM.70.4.2429-2436.2004;  
RA Robertson D.E., Chaplin J.A., Desantis G., Podar M., Madden M.,  
RA Chi E., Richardson T., Milan A., Miller M., Weiner D.P., Wong K.,  
RA McQuaid J., Farwell B., Preston L.A., Tan X., Sneed W.A., Keller M.,  
RA Mathur E., Kretz P.L., Burk M.J., Short J.M.;  
RT "Exploring nitrilase sequence space for enantioselective catalysis.";  
RL Appl. Environ. Microbiol. 70:2429-2436(2004).  
DR EMBL: AY487497; AAR97444.1; -; Genomic DNA.  
DR GO: GO:0018762; Fattyphatic nitrilase activity; IEA.  
DR GO: GO:0016810; F-hydrolyase activity; acting on carbon-nitrog. .; IEA.  
DR GO: GO:0006807; P-nitrogen compound metabolism; IEA.  
DR InterPro: IPR000132; N/CN hydrolase.  
DR Pfam: PF00795; N/CN hydrolase; 1.  
DR PROSITE: PS00263; CN HYDROLASE; 1.  
DR PROSITE: PS00921; NITRIL\_CHT\_2; 1.  
KW Hydrolyase.  
SQ SEQUENCE 346 AA; 37432 MW; 12FE01C35AD463CE CRC64;  
Alignment Scores:  
Pred. No.: 4.47e-100  
Score: 1806.00 Length: 346  
Percent Similarity: 100.0% Matches: 346  
Best Local Similarity: 100.0% Conservative: 0  
Query Match: 96.0% Mismatches: 0  
DB: Indels: 0  
Gaps: 0

US-09-751-299-1 (1-1041) x Q6RMR5\_9Z2ZZ (1-346)

```
QY 1 ATGTCCGAGCCCATGACGAAAGTATCGCGCGCGGCGGTGACAGCCGCGCGGTTCCTC 60
Db 1 Metserglnuprometlnrlystrargllyalalavalglnalalaprovalphenleu 20
QY 61 GATTCGACCGGACAGTGGAGAAAGCGATCGGCTGATCCGACGAGCGGCCAAGCAGAC 120
Db 21 Aspleuasparglnrvalglulysalalaleglyleuileglnlglinalalaleglnasp 40
QY 121 GTGGCGCTGATCGCATTCGCCAGAGACTTGATTCGCCGCTATCCCTTTGGATGAGCTG 180
Db 41 Valargleuilealapheroglnrthrtprleprogllyrprophetripiletrpleu 60
QY 181 GGGCGCGCGGCTTGAGGCGATGCGCTTGCTCGAGCGCTATTCGAGAAATTCGCTGTCGCC 240
Db 61 G1yAlaProAlatrpgllymetArpPheValGlnArglyrPheGlnuanserleuValArg 80
QY 241 GGCAGCAGCAGTGGCAGCGCCCTGCGGATGCGCGCCGCGCAGCGCATGATGTCGTG 300
Db 81 G1ySerlysglntrpglnalaleuAlaAspAlaAlaArgAlaArgAlaArgAlaArgAla 100
QY 301 GCGCGCTATGACGCGCGCGCGCGCGCGCGCATCTATATGCGCGCGCATCTTCGCGCCC 360
Db 101 Alaleglyrsergluarglaleglylserleuutyrmctglglnalalepheglypro 120
QY 361 GATGCGCATGATGCGCGCGCGCGCGCGCGCATGAGCTCAAGCTCAAGCTCAAGCGCAGT 420
Db 121 AspdllyaspLeuilealalealargllyleuylserrothrhlsalaglnuargthrrval 140
QY 421 TTGGCGAGGAGAGACGCGCAGCCATCTTCGCGGTGACATACCGCGCATCGCGCGCTCGC 480
Db 141 PhecllyglulysargllyserhlsleualValhlsasprhralaleglyargleugly 160
QY 481 GCGCTCTGTGCTGGAGCAGCATCCAGCCATTCGAAATAAGCCATGTACGCGCGCGAC 540
Db 161 Alaleucyrcystrpglnhlslelelnproleuserlyrlyalamectryalalaleasp 180
QY 541 GAACAGGTCCACGTCGCGCGCGCGCGCGCATGAGCTCTATGCGCGCGCATGCGCTATGCG 600
Db 181 GlnglnValhlsvalaleasertrproserPheaserleuylrargllymetAlaIyAla 200
QY 601 CTCGACCGGAGGTCAATACCGCGCAGCGCAGATTCACGCGGTGCGAGCGCGGTGCTAC 660
Db 201 Lenglyrprogluvalaenhrhralaleaserclnlelyrlyalaleglnlgllycystyr 220
QY 661 GTGCTGCGCTGTCGCGCGCGCGCGCGCGCGAGATGATGATGATGATGATGATGATGATG 720
Db 221 ValleuAlaSerCySalatrhrvalserProglumectllelysalaleuValaAspThrPro 240
QY 721 GACAGAGAGATGTTCCTCAAGCGCGCGCGCGGTTCGATGATTCGCGCGCGCGCGCGCG 780
Db 241 AspllysglnmetrheleuylsalaleglylPheAlaMetcllepheglyproaspgly 260
QY 781 CCGCGCTGCGCGCGCGCGCGCGCGCGCGAGACCGAAGAGAGACTGTGTGCGCGCATGAC 840
Db 261 ArgAlaleuAlalegluproleuProgluThrglgllyleuValAlaAspIleasp 280
QY 841 CTGCGCATGATCGCGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 900
Db 281 LeugllymetllealeuAlaleuAlaleuAlaleuAlaleuAlaleuAlaleuAlaleu 920
QY 901 GACGTAACGCGCGCTGCTGATGACAGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 960
Db 301 AspValThrArgleuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 320
QY 961 GCAATTGAAACCGGAAACGAGCAGAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGG 1020
Db 321 AlaphegluproclmansnglnuaplysglyaspAlaProAlaleuAlaArgValAlaIagln 340
QY 1021 AGCGCGCGCGCGCGCGAG 1038
|||||
```

Db 341 SerAlaAlaAlaIagln 346

## RESULT 2

Q6RMR4\_9Z2ZZ

ID Q6RMR4\_9Z2ZZ PRELIMINARY; PRT; 341 AA.

AC Q6RMR4

DT 05-JUL-2004 (TEMBLrel. 27, Created)

DT 05-JUL-2004 (TEMBLrel. 27, Last sequence update)

DT 05-JUL-2004 (TEMBLrel. 27, Last annotation update)

DE Nitrilase (EC 3.5.5.7).

GN ORFNames=BD7266;

OS uncultured organism.

OC unclassified: environmental samples.

OX NCBI\_TaxId=155900;

RN [1]

RP NCBILOTIDE SEQUENCE.

RX PubMed=15066841; DOI=10.1128/AEM.70.4.2429-2436.2004;

RA Robertson D.E., Chaplin J.A., Desantis G., Pedar M., Madden M.,

RA Chi E., Richardson T., Milan A., Miller M., Weiner D.P., Wong K.,

RA McQuaid J., Farwell B., Preston L.A., Tan X., Sneed M.A., Keller M.,

RA Mathur E., Kretz P.L., Burk M.J., Short J.M.,

RT "Exploring nitrilase sequence space for enantioselective catalysis.";

RL Appl. Environ. Microbiol. 70:2429-2436 (2004).

DR EMBL; AY487438; AAR97385.1; -; Genomic DNA.

DR GO; GO:0018762; F:aliphatic nitrilase activity; IEA.

DR GO; GO:0016810; F:hydrolyase activity; acting on carbon-nitrog. . .; IEA.

DR InterPro; IPR000132; N/CN hydrolase.

DR InterPro; IPR003010; N/CN hydrolase.

DR Pfam; PF00795; CN\_hydrolase; 1.

DR PROSITE; PS02653; CN\_HYDROLASE; 1.

DR PROSITE; PS00921; NITRIL\_CHT\_2; 1.

KW Hydrolase.

SQ SEQUENCE 341 AA; 36918 MW; 23AC5387FAC1FDEE CRC64;

Alignment Scores:

Pred. No. 1 44e-79 Length: 341

Score: 1464.00 Matches: 272

Percent Similarity: 90.1% Conservative: 28

Best Local Similarity: 81.7% Mismatches: 33

Query Match: 77.8% Indels: 0

DB: 2 Gaps: 0

US-09-751-299-1 (1-1041) x Q6RMR4\_9Z2ZZ (1-341)

```
QY 1 ATGTCCGAGCCCATGACGAAAGTATCGCGCGCGGCGGTGACAGCCGCGCGGTTCCTC 60
Db 1 MetleuserProvalthrlntrglnrlyrargllyalalavalglnalalaleproserPheleu 20
QY 61 GATTCGACCGGACAGTGGAGAAAGCGATCGGCTGATCCGACGAGCGGCCAAGCAGAC 120
Db 21 Aspleuasparglnrvalglulysrhrillealalellelglhlnalaleglulnasp 40
QY 121 GTGGCGCTGATGCAATTCGCCAGAGACTTGATTCGCCGCTATCCCTTTGGATGATGCTG 180
Db 41 ValArgleuileAlaIapheroglnrthrtprleprogllyrproleutripiletrpleu 60
QY 181 GGGCGCGCGGCTTGAGGCGATGCGCTTGCTCGAGCGCTATTCGAGAAATTCGCTGTCGCC 240
Db 61 GlyserProAlatrpgllymetArpPheValGlnArglyrPheGlnuanserleuValArg 80
QY 241 GGCAGCAGCAGTGGCAGCGCCCTGCGGATGCGCGCCGCGCAGCGCATGATGTCGTG 300
Db 81 GlyserlysglntrpAshAlaIalealaleaspAlaAlaArgAlaArgAlaArgAlaArg 320
QY 301 GCGCGCTATGACGAGCGCGCGCGCGCGCGCGCATCTATATGCGCGCGCATCTTCGCGCCC 360
Db 101 ValGlyPheSerGlnArgAlaleglylserleuutyrmctglglnalalepheglypro 120
QY 361 GATGCGCATGATGCGCGCGCGCGCGCGCGCATGAGCTCAAGCTCAAGCTCAAGCGCAGT 420
Db 121 Glngllygluleuilealalealargllyleuylserrothrhlsalaglnuargthrrval 140
```

```
QY 421 TTCGCGAGGAGAGACGCGAGCCATCTCGCGTGACGATACCGCATCGGCGCTCGGC 480
D 141 PheGlyGluGlyArgGlySerHisLeuAlaValTyrGluThrGlyValGlyArgIleGly 160
QY 481 GCGCTCTGTTGCTGAGGACATCCAGCATTTGTGAAATAGCCCATGTATCGCGCGAC 540
D 161 AlaLeuCySerTrpGluHisIleGlnProLeuSerIleTyrAlaMetTyrAlaAlaAsn 180
QY 541 GAAAGATGCCAGTCGCGCTCGTGGCCGAGCTTCAGCCCTTATCCCGGAGATGCCCTTACG 600
D 181 GluGlnValHisValAlaSerTrpProCyPheSerLeuTyrArgGlyMetAlaTyrAla 200
QY 601 CTCGACCGAGATCAATACCGCGCAAGCAGATCTACGCGGTGAGGCGGCTGCTAC 660
D 201 LeuGlyProGluValAsnThrAlaAlaSerGlnValTyrAlaValGluGlyCysTyr 220
QY 661 GTGCTGCGTGTGTCGCGGACCGTTTCGCGGAGATGATCAAGTATTGTGATACGCC 720
D 221 ValLeuAlaSerCybLeuValIleThrProGluIleLeuValIleAspThrPro 240
QY 721 GACACGAGATGTTCTTCACAGCGCGGCGGCTTTTGCATGATTTTCCGCGCGACGCC 780
D 241 AspGlyGluProLeuLeuAlaGlyGlyGlyPheSerMetIlePheGlyProAspGly 260
QY 781 CGCGCGCTGCGCGAGCGCTCCGCGAGACCGAAGAGGAGCTGTGTGCGCGCATATGAC 840
D 261 ArgAlaLeuAlaGlnProLeuProGluThrGluGluGlyLeuValIleThrAlaGluIleAsp 280
QY 841 CTCGCGCATGATCGGCTTGCGCGAAGCGCGCGCGCATCGCGCGGCACTATTCACGCC 900
D 281 LeuGlyAlaIleAlaLeuAlaValAlaAlaAlaAspProAlaGlyHisTyrAlaArgPro 300
QY 901 GACGTAAACGCGGCTGCTGCTGATGACGTCCGCCCAACCGCTGCTGACCTTGATGCC 960
D 301 AspValThrArgLeuLeuAlaAsnProArgProAlaAlaArgValGluAlaLeuGlyPro 320
QY 961 GCATTGCAACGCAAAACGAGACAGAGGCGAGCGCGCC 999
D 321 ArgPheGluValValGlnSerGluGlnAlaGluProPro 333

RESULT 3
Q6RMS2_92ZZZ PRELIMINARY; PRT; 348 AA.
ID Q6RMS2_92ZZZ
AC Q6RMS2_
DT 05-JUL-2004 (TrEMBLrel. 27, Created)
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DE Nitrilase (EC 3.5.5.7).
GN ORFNames=BD5338;
OS uncultured organism.
OC unclassified; environmental samples.
OX NCBI_TaxId=155900;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX PubMed=15066841; DOI=10.1128/AEM.70.4.2429-2436.2004;
RA Roberton D.B., Chaplin J.A., Desantis G., Podar M., Madden M.,
RA Chai E., Richardson T., Milan A., Miller M., Weiner D.P., Wong K.,
RA McQuaid J., Farwell B., Preston L.A., Tan X., Snead M.A., Keller M.,
RA Mathur E., Kretz P.L., Burk M.J., Short J.M.;
RT "Exploring nitrilase sequence space for enantioselective catalysis.";
RL Appl. Environ. Microbiol. 70:2429-2436(2004).
DR EMBL; AY487430; AAR97377.1; -; Genomic_DNA.
DR GO; GO:0018762; F:aliphatic nitrilase activity; IEA.
DR GO; GO:0016810; F:hydrolase activity; acting on carbon-nitrog. . .; IEA.
DR GO; GO:0006807; P:nitrogen compound metabolism; IEA.
DR InterPro; IPR000132; N/CN hydrolase.
DR InterPro; IPR003010; N1ase/CNhydase.
DR InterPro; IPR002332; P11_GlnB_UMP_S.
DR Pfam; PF00795; CN_hydrolase; 1.
DR PROSITE; PS50263; CN_HYDROLASE; 1.
DR PROSITE; PS00920; NITRIL_CHT_1; 1.
DR PROSITE; PS00921; NITRIL_CHT_2; UNKNOWN_1.
DR PROSITE; PS00496; P11_GlnB_UMP; UNKNOWN_1.
```

```
KM HydroLase.
SQ SEQUENCE 348 AA; 38397 MW; 95F18795D75C01A5 CRC64;
Alignment Scores:
Pred. No.: 2,83e-64 Length: 348
Score: 1209.00 Matches: 230
Percent Similarity: 79.3% Conservative: 42
Best Local Similarity: 67.1% Mismatches: 69
Query Match: 64.3% Indels: 2
DB: 2 Gaps: 2

US-09-751-299-1 (1-1041) x Q6RMS2_92ZZZ (1-348)
QY 10 CCATGACGAG---TATCGCGCGCGCGGTGAGCGCGCGCGGTTCCTGATCTC 66
D 2 ProHisSerIysGlnPheArgValAlaAlaValGlnAlaProValPheLeuAspLeu 21
QY 67 GACGCGACAGTCGAGAAAGCGATCGGCTGATCGAGAGCGCGCAAGCAGTGC 126
D 22 GluGlyAlaIleSerIysGlyIleSerLeuIleGluGluAlaIleSerAsnGlyAlaIys 41
QY 127 CTGATCGCATTTCCAGAGACTTGATTCGCGCTATTCCTTTTGATATGCTGCGCGC 186
D 42 LeuIleAlaPheProGluThrTrpIleProGlyTyrProTrpTrpIleTrpLeuAspSer 61
QY 187 CGGCTTTGGGGCATCGGCTTCGTCCAGCGCTATTCGAGAAATTCGTCGCGCGAGC 246
D 62 ProAlaTrpAlaIleSerIysGlnPheArgValGlnArgTyrPheAspAsnSerLeuMetLeuGlySer 81
QY 247 AAGCAGTGGACGAGCGCTTGCGGATGCGCGCGCGCGCGCATGATGTCGTGCGCGC 306
D 82 GluGlnAlaIysArgMetAsnGlnAlaAlaAlaAsnAsnIysIleTyrValIleMetGly 101
QY 307 TATAGCAGCGCGCGCGCGCGAGCTTATATAGGCCAGCGCATCTTGCGCGCGATGCC 366
D 102 TyrSerGluArgSerGlyIleSerLeuTyrMetGlyGlnSerIleIleAsnAspIysGly 121
QY 367 GATCTATGCCCGCGCGCGCGCGCATGAGCTTACGCCATCGGAGCGGACCGTGTGGC 426
D 122 GluThrIlePheThrArgArgIysLeuIysProThrIleValGluAlaGluThrValPheGly 141
QY 427 GAGGAGACGCGGACCCATCTCGCGGTGACAGATACCGCATCGGCGCGCTCGCGCGCTC 486
D 142 GluGlyAspGlySerHisLeuCybValMetAspThrGluIleGlyArgValGlyAlaMet 161
QY 487 TGTGCTGAGGACATCCAGCATTTGTCAAAATAGCCCATGTACGCGCGCGAGCAACG 546
D 162 CySerTrpGluHisLeuGlnProLeuSerIleTyrAlaMetTyrSerGlnAspGluGln 181
QY 547 GTCCAGCTCGCGTCGTGCGCGGAGCTTACGCTTATGCGCGGATGCTTATGCGTGGGA 606
D 182 IleHisIleAlaSerTrpProSerPheSerLeuTyrGlyAlaAlaTyrAlaLeuGly 201
QY 607 CCGAGGTCAATACCGCGCGAGCAGATCTACGCGGTGACGAGGCGGCTGATGCTGCTG 666
D 202 ProGluLeuAsnAlaAlaAlaSerGlnMetTyrAlaAlaGluGlyGlnCybPheValLeu 221
QY 667 GCGTGTGCGCGACCGTTTCGCGGAGATGATCAAGTATTGTGTAGTACGCCCGAGAAC 726
D 222 AlaProCybAlaThrValSerIysGluMetIleGluMetLeuIleAspAspProArgIys 241
QY 727 GAGATGTTCTCAAGCGCGCGCGGCTTTTGCATGATTTTCCGCGCGCGGCGCGCGCC 786
D 242 GluProLeuLeuLeuGluGlyGlyGlyPheThrMetIleTyrGlyProAspGlyArgPro 261
QY 787 CTGCGCGAGCGGCTCCGCGAGACCGAAGAGAGACTGTGTGCTCCGATATGACTGCGC 846
D 262 LeuAlaIysProLeuProGluAsnGluGluGlyLeuLeuTyrAlaAspIleAspLeuGly 281
QY 847 ATGATCGGTTGGCCAAAGCGCGCGCGATCCGCGCGGCGCATTTTCAAGCGCGGACGTA 906
D 282 MetIleSerMetAlaIysAlaAlaAlaAspProAlaGlyHisTyrAlaArgProAspVal 301
```

```

QY 907 ACGGCGCTGCTGATGACGTCGCGCCCAACGCTGTCACGCTTGATGATCCGACATTC 966
Db 302 ThrhgluLeuLeuPheAsnSerAlaProAlaAsnAlaGluTyrTleAsnProAlaSer 321
QY 967 GAACCGCAAAAGAGAGACAGAGCGCGCGCGCGCTGAGTGGTGGGAAAGCGCC 1026
Db 322 GlyPro--ThrGluSerLeuLeuAspMetGlyLysMetGlnMetGluAlaGluGln 340
QY 1027 GCCCGCGCG 1035
Db 341 LysAlaAla 343

RESULT 4
Q6RMF9_9ZZZZ PRELIMINARY; PRT; 353 AA.
ID Q6RMF9_9ZZZZ
AC Q6RMF9_
DT 05-JUL-2004 (TReMBLrel. 27, Created)
DT 05-JUL-2004 (TReMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TReMBLrel. 27, Last annotation update)
DE Nitrilase (EC 3.5.5.7).
GN ORFNames=BD7753;
OS uncultured organism.
OC unclassified; environmental samples.
OX NCBI_TaxID=155900;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX PubMed=15066841; DOI=10.1128/AEM.70.4.2429-2436.2004;
RA Robertsson D.E., Chaplin J.A., Desantis G., Podar M., Madden M.,
RA Chi E., Richardson T., Milan A., Miller M., Weiner D.P., Wong K.,
RA McQuaid U., Farwell B., Preston L.A., Tan X., Sneed M.A., Keller M.,
RA Mathur E., Kretz P.L., Burk M.J., Short J.M.,
RT "Exploring nitrilase sequence space for enantioselective catalysis.";
RL Appl. Environ. Microbiol. 70:2429-2436(2004).
DR EMBL; AY487543; AAR97490.1; -; Genomic DNA.
DR GO; GO:0018762; F:aliphatic nitrilase activity; IEA.
DR GO; GO:0016810; F:hydrolase activity, acting on carbon-nitrog. . .; IEA.
DR GO; GO:0006807; P:nitrogen compound metabolism; IEA.
DR InterPro: IPR001032; N/CN hydrolase.
DR InterPro: IPR003010; N/CN hydrolase.
DR Pfam; PF00795; CN_hydrolase; 1.
DR PROSITE; PS50263; CN_HYDROLASE; 1.
DR PROSITE; PS00920; NITRIL_CHT_1; 1.
DR PROSITE; PS00921; NITRIL_CHT_2; 1.
KM Hydrolase.
SQ SEQUENCE 353 AA; 38029 MW; 997D13779FC1326 CRC64;

Alignment Scores:
Pred. No.: 1.23e-61 Length: 353
Score: 1165.00 Matches: 229
Percent Similarity: 78.8% Conservative: 31
Best Local Similarity: 69.4% Mismatches: 66
Query Match: 61.9% Indels: 4
DB: 2 Gaps: 3

US-09-751-299-1 (1-1041) x Q6RMF9_9ZZZZ (1-353)
QY 19 AAGTATCGCGCGCGCGCGCGCGCGCGCGTCTCGATCTCGACCGACAGTC 78
Db 7 LysTyrLeuValAlaAlaValAlaAlaAlaProAlaPheLeuAspAlaSerVal 26
QY 79 GAGAAAGCATCGCGCTGATCGAGACGCGCCAGACAGACGTCGCGCTGATCGCATTC 138
Db 27 GluLysAlaValAlaPheLeuAlaGluAlaAlaGluAlaValAlaAlaPhe 46
QY 139 CCAAGACTTGATCCCGGCTATCCCTTGGATGATGCTGGCGCGCGCTGGGGC 198
Db 47 ProGluThrTrpLeuProGlyTyrProTrpTrpLeuGlyAlaProAlaTrpAla 66
QY 199 ---ATGGC---TTCCGTCACGCGTATTTGAGAATTGCTGCGCGCGACGACAG 252
Db 67 IleMetArgGlyPheValSerArgTyrPheAspAsnSerLeuSerTyrAspSerProGln 86
QY 253 TGGCAGGCGCTGGCGGATGCGCGCGCGCGCACGCGATGATGTCGTCGCGCTATAGC 312

```

```

Db 87 AlaGluLysLeuArgAlaAlaAlaLysArgAsnLysMetValValAlaLeuGlyLeuSer 106
QY 313 GAGCGCGCGCGCGACGCTCTATATGGCGCCGATCTTCGAGCCCGGATGGGATCTG 372
Db 107 GluArgAspGlyGlySerLeuTyrTleAlaGlnTrpLeuTleIleGlyProAspGlyGluThr 126
QY 373 ATGCGCGCGCGCGCAAGCTTCAAGCTTACCGATGCGGAGCGCACCGTGTTCGAGAGGA 432
Db 127 IleAlaLysArgAlaGlyLysLeuLysProThrAlaAlaGluArgThrValPheGlyGluGly 146
QY 433 GACGCGACGATCTCGCGGATGACAGATACCGGATGCGGCGCGCGCTGCTGTTC 492
Db 147 AspGlySerHisLeuAlaValHisGluLeuAspValGlyArgLeuGlyAlaLeuCyS 166
QY 493 TGGAGCACAATCAAGCCATTTGTGAATATACGCGATGACGCGCGACGACAGATTCAC 552
Db 167 TrpGluHisLeuGlnProLeuSerTyrTyrAlaMetTyrAlaGlnGlnGluValHis 186
QY 553 GTCGCGCTGCGCGGAGCTTCAAGCTTATCGCGGATGCGGCGCGCTGCGGACGAG 612
Db 187 ValAlaAlaTrpProSerPheSerLeuTyrAspProPheAlaHisAlaLeuGlyAlaGlu 206
QY 613 GTCATATACGCGCGCAAGCCAGATCTACGCGGTGAGAGCGCGGCTGTAAGTGTGCGCTCG 672
Db 207 ValAsnAsnAlaAlaSerTyrTyrAlaValAlaGluGlySerCySAspValIleAlaPro 226
QY 673 TGGCGGACGCTTTCGCGGAGATGATCAAGTATGTTGTGATCGCGCGCAAGAGATG 732
Db 227 CysAlaThrValSerGlnAlaMetIleAspGluLeuCySAspThrProGluLysHisGln 246
QY 733 TTCCTGAGCGCGCGCGGTTTTCGATGATTTTCGAGCGCGCGCGCGCGCTGAGCC 792
Db 247 PheLeuHisAlaGlyGlyGlyPheAlaValIleTyrGlyProAspGlyAlaProLeuAla 266
QY 793 GACCGCGCTCGCGGACCGACAGAGGAGCTGCTGCGCGGATTCGACTTCGCGCATGATC 852
Db 267 AlaProLeuProAspLysGluGlyLeuLeuTyrAlaAspIleAspLeuGlyMetIle 286
QY 853 GCGTTGGCCAAAGCGCGCGCGATCCGCGGCGCATATTCACGCGCGACGATACGCGG 912
Db 287 SerValAlaLysValAlaAlaAspProAlaGlnHisTyrAlaArgProAspValThrArg 306
QY 913 CTGCTGCTGATGACGTCGCGCGCCAAAGCGCTGTCAGCTTGATCCGATTCGAACCG 972
Db 307 LeuLeuPheAsnAsnArgProGlyTyrArgValAlaGluThrMet-----AlaLeuProIle 324
QY 973 CAAGAAGAGACAGAGCGCGCGCGCGCG 1002
Db 325 AspAlaGluThrLysAlaGluAlaProAla 334

RESULT 5
Q6RMF9_9ZZZZ PRELIMINARY; PRT; 353 AA.
ID Q6RMF9_9ZZZZ
AC Q6RMF9_
DT 05-JUL-2004 (TReMBLrel. 27, Created)
DT 05-JUL-2004 (TReMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TReMBLrel. 27, Last annotation update)
DE Nitrilase (EC 3.5.5.7).
GN ORFNames=BD5217;
OS uncultured organism.
OC unclassified; environmental samples.
OX NCBI_TaxID=155900;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX PubMed=15066841; DOI=10.1128/AEM.70.4.2429-2436.2004;
RA Robertsson D.E., Chaplin J.A., Desantis G., Podar M., Madden M.,
RA Chi E., Richardson T., Milan A., Miller M., Weiner D.P., Wong K.,
RA McQuaid U., Farwell B., Preston L.A., Tan X., Sneed M.A., Keller M.,
RA Mathur E., Kretz P.L., Burk M.J., Short J.M.,
RT "Exploring nitrilase sequence space for enantioselective catalysis.";
RL Appl. Environ. Microbiol. 70:2429-2436(2004).
DR EMBL; AY487552; AAR97499.1; -, Genomic DNA.

```







Db 207 ValAsnAsnAlaAserIyAlleTyAlaValAlGluGlySerCySPheValIleAlaPro 226  
Qy 673 TGGCGACCGTTTGCCGCGAGATGATCAAGGTATTGGTATACGCCGACAGAGATG 732  
Db 227 CysAlaThrValSerGlnAlaMetIleAspGluLeuCyAspSerProGluLysIleGln 246  
Qy 733 TTCCTCAAGCGCGCGGGGCGTTTGCCATGATTTTCGGGCGCGGCGCGCGCGTGGCC 792  
Db 247 PheLeuHisValAlGlyGlyPheAlaValIleTyAlaGlyProAspGlyAlaProLeuCy 266  
Qy 793 GAGCGGCTCCCGAGACCGAAGAGGAGCTGTGTGCGCGATATCGACTTCGCGATATC 852  
Db 267 LysProLeuAlaProAspGlnGluGlyLeuLeuTyAlaAlaPheLeuGlyMetIle 286  
Qy 853 GCGTTGGCCAAAGCGCGCGCGCGATCCGCGCGGCGGCGCATTTACCGCGCGCGTAAACGG 912  
Db 287 SerValAlaLysAlaAlaAlaAspProAlaGlyHisTyAlaArgProAspValThrArg 306  
Qy 913 CTGCTGCTGCATCGACGTCGCGCGCCGACGCTGTCAGCTTGATGCCGCAATTGGACCG 972  
Db 307 LeuLeuPheAsnAsnArgProGlyAsnArgValGluThrLeuAlaLeuProValAspGln 326  
Qy 973 CAACAACGAG--GACAAAGCGACACGCGCGCGCTGCGCTGTGCGCGAAGCGCC-- 1026  
Db 327 GluAlaGluAlaGlyAlaGlyGlyLysProAlaProLysSerProSerValAlaAlaPhe 346  
Qy 1027 -----GCCGCGCGCGAG 1038  
Db 347 ThrLeuThrGlnAlaAlaAlaGlu 354  
RESULT 8  
Q706Q8\_PSEPU PRELIMINARY; PRT; 331 AA.  
ID Q706Q8; AC Q706Q8;  
DT 05-JUN-2004 (TREMBlrel. 27, Created)  
DT 05-JUN-2004 (TREMBlrel. 27, Last sequence update)  
DT 05-JUN-2004 (TREMBlrel. 27, Last annotation update)  
DS Predicted amidohydrolase/nitrilase.  
OS Pseudomonas putida.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
OC Pseudomonadaceae; Pseudomonas.  
NCBI\_TaxID=303;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=strain RR21;  
RX MEDLINE=99008981; PubMed=9791097;  
RA Ravehn R., Studer S., Zander A.J.B., van der Meer J.R.;  
RT "Int-B13, an unusual site-specific recombinase of the bacteriophage P4  
RT kb c1c-element of Pseudomonas sp. strain B13.";  
RL J. Bacteriol. 180:5505-5514 (1998).  
[2]  
RP NUCLEOTIDE SEQUENCE.  
RC STRAIN=strain RR21;  
RX MEDLINE=22708816; PubMed=12823813;  
RA DOI=10.1046/j.1365-2958.2003.03548.x;  
RA Senthil V., Zehnder A.J.B., van der Meer J.R.;  
RT "Characterization of two alternative promoters and a transcription  
RT regulator for integrase expression in the c1c catabolic genomic island  
RT of Pseudomonas sp. strain B13.";  
RL Mol. Microbiol. 49:93-104 (2003).  
[3]  
RN NUCLEOTIDE SEQUENCE.  
RC STRAIN=strain RR21;  
RA Galliard M., Werlen C., Vallaeys T., Vorhoeiter F.J., Puehler A.,  
RA van der Meer J.R.;  
RL Submitted (DEC-2003) to the EMBL/Genbank/DBJ databases.  
DR EMBL; AJ617740; CAB92892.1; -; Genomic DNA.  
DR GO; GO:0016810; F:hydrolase activity, acting on carbon-nitrog. . .; IBA.  
DR InterPro; IPR000132; N/CN hydrolase.  
DR InterPro; IPR001010; NClase/CNhydrolase.  
DR Pfam; PF00795; CN\_hydrolase; 1.

DR PROSITE; PS50263; CN HYDROLASE; 1.  
DR PROSITE; PS00920; NITRIL\_CHT\_1; 1.  
DR PROSITE; PS00921; NITRIL\_CHT\_2; UNKNOWN\_1.  
KW Hydrolase.  
SQ SEQUENCE 331 AA; 36554 MW; DEF59D03874D9187 CRC64;  
Alignment Scores:  
Pred. No.: 2 33e-59 length: 331  
Score: 1127.00 Matches: 206  
Percent Similarity: 79.4% Conservative: 45  
Best Local Similarity: 65.2% Mismatches: 65  
Query Match: 59.9% Indels: 0  
Gaps: 0  
US-09-751-299-1 (1-1041) x Q706Q8\_PSEPU (1-331)  
Qy 13 ATGACGAGTATCGCGCGCGCGGCGGTACAGCGCGCGCGGTGTTCTCGATTCGACCGC 72  
Db 6 IleAlaLysTyAlaValAlaAlaValAlGlnAlaAlaProGluPheLeuAsnLeuAspLys 25  
Qy 73 ACAGTCGAGAAAGCATCGCGCTGATTCAGACGCGCGCGCAAGACAGCTCCGCTGATC 132  
Db 26 GlyValGluLysAlaValaLArgLeuIleGluGluAlaAlaLysAsnGlyAlaLysLeuIle 45  
Qy 133 GCATTCGCGAGACTTGATTCGCGGCTATCCCTTTGATATGATGCTGGCGCGCGGCT 192  
Db 46 AlaPheProGluValThrPheProGlyTyProThrPheLeuAspSerProAla 65  
Qy 193 TGGGGCATGCGCTTCGTCAGCGGTATTCAGAAATTCGTCGTCGCGCGACGACGAG 252  
Db 66 TrpGlyMetGlnPheValGlnArgHisPheGluAsnAlaLeuLeuValGlySerProGln 85  
Qy 253 TGGCAGGCGCTGGCGGATGCGCGCGCGCGCGCGCATGCAATGTCGTGCGCGCTATAC 312  
Db 86 TrpGluArgLeuCysAlaAlaAlaAlaGluHisArgGlyLeuPheValLeuGlyPheCys 105  
Qy 313 GAGCGCGCGCGCGCGCGCTATATGCGCGCGCATTCGCGCGCGCGCGCGATGCG 372  
Db 106 GluArgGlnAspGlyThrLeuTyAlaAlaGlnAlaIleLeuAspAspGluLysArgVal 125  
Qy 373 ATCGCGCGCGCGCGCGCATGACCTTACGCCATGCGAGCGCACCGTGTTCGCGAGGGA 432  
Db 126 ValSerThrArgArgLysLeuLysProThrHisAlaGluArgThrValTyGlyGluGly 145  
Qy 433 GACGCGACCATTCGCGGTGCGACGATACCGCATCGCGCGCTCGCGCGCTGTGTC 492  
Db 146 AspGlySerHisLeuSerValHisGlnThrSerIleGlyArgIleGlyAlaLeuSerCys 165  
Qy 493 TGGGAGCATTCGACCATGTCGAATATGACCATGTAACCGCGCGCGCAAGACGATCAC 552  
Db 166 AlaGluHisIleGlnProLeuSerLysTyAlaMetTyIserGlnAsnGluGlnIleHis 185  
Qy 553 GTCGCTGTCGCGCGACCTTCAGCCTTATCGCGGCGCATGCGCTTATGCGTCGACCGGAG 612  
Db 186 ValAlaAlaThrProSerPheSerValTyArgGlyAlaAlaPheGlnLeuSerProGlu 205  
Qy 613 GTCATATCGCGCGACCGACGATTCAGCGGTGCGAGCGGCGGTCTACGTGCGCTCG 672  
Db 206 AlaAsnLeuAlaAlaSerGlnValTyAlaLeuGluGlyCysTyAlaLeuAlaPro 225  
Qy 673 TGGCGGACGTTTTCGCGGAGATGATCAAGGTATTGGTGTGATACCGCGCAAGAGATG 732  
Db 226 CysAlaLeuValSerLysGluMetLeuLeuAlaAspThrProGlnLysArgGln 245  
Qy 733 TTCCTCAAGCGCGCGCGGTTTGCCATGATTTTCGGGCGCGGCGCGCGCTTGACC 792  
Db 246 LeuLeuLeuGluGlyGlyPheAlaGlnIlePheHisProAspAlaLysProLeuCys 265  
Qy 793 GAGCGGCTCCCGAGACCGAAGAGGAGCTGTGTGCGCGATATCGACTTCGCGATATC 852  
Db 266 GluProPheProGluThrGlnGluGlyLeuLeuTyAlaAspValaAspLeuGlyPheIle 285  
Qy 853 GCGTTGGCCAAAGCGCGCGCGATCCGCGCGGCGCATTCACGCGCGCGCGCATGACCGCG 912

```

Db      286 GlyValAlaValAspAlaValTyrAspProThrGlyHisTyrSerArgProAspValValArg 305
Qy      913 CTGCTGTCGATCGACGTCGCCGCCAACGCCGTCGTCGATGATGCC 960
Db      306 LeuLeuTrpAsnSerLysProAlaThrArgValHisSerPheGlnAla 321

RESULT 9
Q6RWS5_9ZZZZ PRELIMINARY; PRT; 345 AA.
ID      Q6RWS5_9ZZZZ
AC      Q6RWS5;
DT      05-JUL-2004 (TREMBLrel. 27, Created)
DT      05-JUL-2004 (TREMBLrel. 27, Last sequence update)
DT      05-JUL-2004 (TREMBLrel. 27, Last annotation update)
DE      Nitrilase (EC 3.5.5.7).
GN      ORFNames=BD7611;
OS      uncultured organism.
OC      unclassified; environmental samples.
OX      NCBI_TaxID=155900;
RN      [1]
RP      NUCLEOTIDE SEQUENCE.
RX      PubMed=15066841; DOI=10.1128/AEM.70.4.2429-2436.2004;
RA      Chai E., Richardson T., Milan A., Miller M., Weiner D.P., Wong K.,
RA      McQuaid J., Farwell B., Preston L.A., Tan X., Sneed M.A., Keller M.,
RA      Machur E., Kretz P.L., Burk M.J., Short J.M.;
RT      "Exploring nitrilase sequence space for enantioselective catalysis.";
RL      Appl. Environ. Microbiol. 70:2429-2436(2004).
DR      EMBL; AY487427; AAR97374.1; -; Genomic DNA.
DR      GO; GO:0018762; F:aliphatic nitrilase activity; IEA.
DR      GO; GO:0006807; P:nitrilase activity, acting on carbon-nitrog. . .; IEA.
DR      InterPro; IPR000132; N/CN_hydrolase.
DR      InterPro; IPR003010; Ntlase/CNhydase.
DR      Pfam; PF00795; CN_hydrolase; 1.
DR      PROSITE; PS00263; CN_HYDROLASE; 1.
DR      PROSITE; PS00920; NITRIL_CHT_1; 1.
DR      PROSITE; PS00921; NITRIL_CHT_2; 1.
KM      Hydrolase.
SQ      SEQUENCE 345 AA; 37242 MW; 5F8A3CF530D2F1D CRC64;

Alignment Scores:
Pred. No.: 3,09e-59 Length: 345
Score: 1125.00 Matches: 212
Percent Similarity: 77.1% Conservative: 44
Best Local Similarity: 63.9% Mismatches: 72
Query Match: 59.8% Indels: 4
DB: Gaps: 1

US-09-751-299-1 (1-1041) x Q6RWS5_9ZZZZ (1-345)
Qy      19 AAGTATGCGCGCGCGCGCGCGCGCGCGCGGTCGATCTCGAACCGCAGTC 78
Db      7 ArgTyrArgValAlaAlaValAlaGlnAlaAlaProGlnPheLeuAsnLeuGlnAlaThrVal 26
Qy      79 GAGAAACGATCGCGCTGATCGACGACGCGCCAAAGAGAGAGCGTGCCTGATGCATTC 138
Db      27 AspLysThrIleAlaLeuIleGlnGlnAlaAlaArgGlyAlaAsnSerLysIleAlaPhe 46
Qy      139 CCAAGACTGATTCGCGGCTATCCGCTTTCGATGATGCGGCGCGCGCGCTTGCGGC 198
Db      47 ProGlnThrTrpIleProGlyTyrProTrpPheAlaTrpLeuGlyAlaProIleTrpGly 66
Qy      199 ATGGCGCTTCGACCGCTATTTGAGAAATTCGTCGTCGCGCGCAGCAAGCAGTGGCAG 258
Db      67 MetLysPheIleGlnAlaTyrHisAspAsnSerMetValIleAspGlyAlaGlnPheGlu 86
Qy      259 GCCCTGGCGGATGGCGCCCGCGCAGCAGCATGATCTGTCGCGCGCTATGAGAGCC 318
Db      87 ArgIleAlaGlnAlaAlaSerArgCysAsnIleThrValValLeuGlyPheSerGluLys 106
Qy      319 GCGGCGCGAGCCTCTATATGCGCGCAGGAGCATCTTCGCGCCGATGCGATCTGATGCC 378

```

```

Db      107 AspAlaGlySerLeuTyrIleAlaGlnAlaIleLeuSerProGlnGlyLysThrIleAla 126
Qy      379 GCGGCGCGCAAGCTCAAGCTTACCCTACCGAGAGCGACCGTGTTCGCGAGGAGAGCGC 438
Db      127 ThrArgTrpGlyLeuLysProThrHisValGlnArgAlaIlePheGlnGlyGlnAlaAspGly 146
Qy      439 AGCCATTCGCGGTGACCGATACCGCCATCGGCGCCTCGCGCGCTCTGTGTCGGAG 498
Db      147 SerAspLeuAlaValHisAspThrLysLeuGlyArgValGlyAlaLeuCysCysTrpGlu 166
Qy      499 CACATCCAGCATTTGCGAAATAGCCCATTCAGCGCCGCGCAGCAAGCAGTCCAGTCGCG 558
Db      167 HisLeuGlnProLeuSerLysTyrAlaMetTyrAlaGlnGlnGlnGlnValHisIleAla 186
Qy      559 TCGTGGCGGAGCTTCAGCCTCTATTCGGGCGATGCGCCTATTCGCTCGGACCGAGTCAAT 618
Db      187 AlaTrpProSerPheSerLeuTyrValAspAlaAlaTyrAlaLeuGlyProGlnValAsn 206
Qy      619 ACCGCGCAAGCAGATCTACGCGGTGAGAGGCGGCTACGTCGTGCGTGGCGCG 678
Db      207 AsnAlaAlaSerArgLeuTyrAlaValGlnGlyGlnCysPheValValAlaProCysAla 226
Qy      679 ACCGTTTCGCGGAGATGATCAAGTATTCGTGATACGCCCAAGAGATGTTCTTC 728
Db      227 ThrValSerGlnLysMetIleAspMetLeuCysGlnThrProGlnGlnGlnAlaLeuLeu 246
Qy      739 AAGCGCGCGCGGATTTGTCATGATTTTCGCGCGCGCGCGCGCGCGCGCGCGAGCG 728
Db      247 LysProGlyLysIleGlnHisAlaGlnIleTyrGlyProAspGlyArgSerLeuAlaAspPro 266
Qy      799 CTCGCCGAGACCGAAGAGGAGCTGTGTCGCCGATTCGACTTCGCGATGATCGCGTG 858
Db      267 LeuProProAspAlaGlnGlyLeuTyrAlaAspIleAspLeuAlaAlaIleThrLeu 286
Qy      859 GCCAAGCGCGCGCGCGATCCGCGGCGCACTATTCACGCGCGCAAGCAAGCGCGCTG 918
Db      287 AlaLysAlaAlaAlaAspProAlaGlnIleTyrSerArgProAspValThrGlnLeuLeu 306
Qy      919 CTGATTCGACGTCGCGCCCAAGCGCTGTCACGCTTGATGCGCATTCGACCGCAAAAC 978
Db      307 LeuAspArgAsnProLysProArgValAlaHisAla-----LysProGlnGln 322
Qy      979 GAGAGCAAGGCGACGCGCGCGCGCGCGCGCGTGTG 1014
Db      323 SerAlaAsnAsnSerProGlyMetArgAlaVal 334

RESULT 10
Q6RWS3_9ZZZZ PRELIMINARY; PRT; 334 AA.
ID      Q6RWS3_9ZZZZ
AC      Q6RWS3;
DT      05-JUL-2004 (TREMBLrel. 27, Created)
DT      05-JUL-2004 (TREMBLrel. 27, Last sequence update)
DT      05-JUL-2004 (TREMBLrel. 27, Last annotation update)
DE      Nitrilase (EC 3.5.5.7).
GN      ORFNames=BD7051;
OS      uncultured organism.
OC      unclassified; environmental samples.
OX      NCBI_TaxID=155900;
RN      [1]
RP      NUCLEOTIDE SEQUENCE.
RX      PubMed=15066841; DOI=10.1128/AEM.70.4.2429-2436.2004;
RA      Robertson D.E., Chaplin J.A., Desantis G., Podar M., Madden M.,
RA      Chai E., Richardson T., Milan A., Miller M., Weiner D.P., Wong K.,
RA      McQuaid J., Farwell B., Preston L.A., Tan X., Sneed M.A., Keller M.,
RA      Machur E., Kretz P.L., Burk M.J., Short J.M.;
RT      "Exploring nitrilase sequence space for enantioselective catalysis.";
RL      Appl. Environ. Microbiol. 70:2429-2436(2004).
DR      EMBL; AY487559; AAR97506.1; -; Genomic DNA.
DR      GO; GO:0018762; F:aliphatic nitrilase activity; IEA.
DR      GO; GO:0006807; P:nitrilase activity, acting on carbon-nitrog. . .; IEA.
DR      InterPro; IPR000132; N/CN_hydrolase.
DR      InterPro; IPR003010; Ntlase/CNhydase.

```



```
Db      87 AlaGluYsIleArgLyAlaValIysArgAlaValIleuThrAlaValIleGlyValSer 106
Qy      313 GAGCGCGCGGCGGAGCCTCTATATGGCCAGGCACTTCGCGCCGATGGCGATCTG 372
Db      107 GluArgAspGlyGlySerIleuYrIleGlyInTrIleuIleGlyProAspGlyGluThr 126
Qy      373 ATCCGCGCGCGCGGAGCACTTACCGATACCGATGGGAGCCACCGTTTGGCGAGGA 432
Db      127 IleAlaYsArgValYsIleuThrProThrIleAlaGluArgThrValPheGlyGluGly 146
Qy      433 GACGCGACCATCTTCGCGGTGACGATACCGCC--ATCGGCGCGCTTCGCGCGCTCTGT 489
Db      147 AspGlySerAspIleuAlaValIleAspArgAlaAspValGlyArgLeuGlyAlaMetCys 166
Qy      490 TGCTGGGAGCAATCCAGCCATTCGCAATACCGCATATACCGCGCGGAGCAAGGATC 549
Db      167 CysTrpGluHsIleuGlnProIleuSerIyTrAlaMetYrAlaGlnAsnGluGlnVal 186
Qy      550 CACGTGCGGTGCGCGCGGATTCAGCTCTATCGCGGATGCGCTATGCGATCGAGACG 609
Db      187 HisValGlyAlaTrpProSerPheSerIleuYrAspProPheAlaHisAlaLeuGlyTrp 206
Qy      610 GAGGTCAATACCGCGCGGAGCAATCTACCGCGGTGAGGCGCGCTGTCTACGTCTGCG 669
Db      207 GluValAsnAsnAlaAlaSerIleValYrAlaValGluGlySerCysPhePheLeuGly 226
Qy      670 TCGTGGCGGACGTTTCCGCGGAGATGATCAAGGATATGTGATGATACCGCGGAGAG 729
Db      227 ProCysAlaValValSerGlnAlaMetIleAspGluIleuGlyAspSerProGluYHis 246
Qy      730 ATGTTCTCAAGCGCGCGCGGTTCGCAATATTCGGGCGCGGAGCGCGCGCTGTG 789
Db      247 AlaPheIleuHsAlaGlyGlyIleAlaValIleYrGlyProAspGlySerSerIleu 266
Qy      790 GCCGCGCGCTCCGCGAGACCGCAAGAGGATGCTGTCGCGGATATCTGAGCTCGAGATG 849
Db      267 AlaAspYsIleuProProAspGlnGluGlyIleuYrAlaAspIleAspLeuGlyMet 286
Qy      850 ATCGCGTTGGCGGAGCGCGCGCGATTCGCGCGGCGCACTATTCAGCGCGGAGCTAAG 909
Db      287 IleGlyAlaValAlaYsAsnAlaAlaAspProIaGlyHsIySerArgProAspValThr 306
Qy      910 CGGCTGCTGATGATCGACGTCCGCGCCCAACGCGCTC 945
Db      307 ArgIleuIleuAsnThrSerArgAlaAsnArgVal 318

RESULT 12
Q6RMG6_9222Z PRELIMINARY; PRT; 338 AA.
ID Q6RMG6_9222Z PRELIMINARY; PRT; 338 AA.
AC Q6RMG6;
DT 05-JUL-2004 (TREMBlrel. 27, Created)
DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)
DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)
DE Nitrilase (EC 3.5.5.7).
GN ORFNames=BD5279;
OS uncultured organism.
OC unclassified; environmental samples.
OX NCBI_TaxID=155900;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX PubMed=1506841; DOI=10.1128/AEM.70.4.2429-2436.2004; Madden M.,
RA Roberton D.E., Chaplin J.A., Desantis G., Podar M.,
RA Chi E., Richardson T., Milan A., Miller M., Weiner D.P., Wong K.,
RA McQuaid J., Farwell B., Preston L.A., Tan X., Sneed M.A., Keller M.,
RA Machur E., Kreza P.L., Burk M.J., Short J.M.;
RA "Exploring nitrilase sequence space for enantioselective catalysis.";
RT Appl. Environ. Microbiol. 70:2429-2436(2004).
RL EMBL: AY487536; AAR97483.1; -; Genomic DNA.
DR GO: GO:0018762; F:aliphatic nitrilase activity; IEA.
DR GO: GO:0018610; F:hydrolyase activity; acting on carbon-nitrog. .; IEA.
DR GO: GO:0006807; P:nitrogen compound metabolism; IEA.
DR InterPro: IPR000132; N/CN_hydrolase.
```

```
DR InterPro: IPR003010; Ntlase/CNhydase.
DR Pfam: PF00795; CN_hydrolase; 1.
DR PROSITE: PS0263; CN_HYDROLASE; 1.
DR PROSITE: PS00920; NITRIL_CHT_1; 1.
DR PROSITE: PS00921; NITRIL_CHT_2; 1.
KW Hydrolase.
SQ SEQUENCE 338 AA; 36483 MW; B9EBD7FC216498BD CRC64;
Alignment Scores:
Pred. No.: 3,77e-56 Length: 338
Score: 1073.50 Matches: 213
Percent Similarity: 77.9% Conservative: 30
Best local Similarity: 68.3% Mismatches: 66
Query Match: 57.1% Indels: 3
DB: Gaps: 2
US-09-751-299-1 (1-1041) x Q6RMG6_9222Z (1-338)
Qy      19 AAGTATCGCGCGCGGCGGTGAGGCGCGCGGTGTTCTGATTCGACCGCAGATC 78
Db      7 LysTyrArgValAlaAlaValAlaGlnAlaAlaProAlaTrpIleuAspIleuAspArgSerIle 26
Qy      79 GAGAAAGCGATCGGCTGATCGAGCAGCGCGCGCAAGACGATGCGCTGATCGATTC 138
Db      27 AspIysAlaIleAlaIleuIleGluGlnAlaAlaAlaAsnGlyAlaArgIleuIleAlaPhe 46
Qy      139 CGAGAGCTTGATTCGCGGCTATCCCTTTGGATATGCTGGGCGCGCGCTTG--- 195
Db      47 ProGluValPheIleProGlyTrpProThrPheIleTrpIleuAspSerProAlaTrpAla 66
Qy      196 --GCAATGCGCTTCCTCGACGCTATTCGAAATTCGCTCGTGGCGGAGCAAGCAG 252
Db      67 IleGlyArgGlyPheValGlnArgTyrPheAspAsnSerIleuAlaTrpAspSerProGln 86
Qy      253 TGCGAGCGCTGCGGATCGCGCGCGCGCGCAACGCGATGATGCTGCGCGGCTATACG 312
Db      87 AlaGluArgIleuArgAlaAlaValAlaArgLyAlaArgLeuThrAlaValIleGlyLeuSer 106
Qy      313 GAGCGCGCGGCGGAGCCTCTATATGGCCAGGCACTTCGCGCCGATGGCGATCTG 372
Db      107 GluArgSerGlyGlySerIleuYrIleGlyInTrIleuIleGlyProAspGlyGluThr 126
Qy      373 ATCGCGCGCGCGGAGCACTTACCGCATATCCGATTCGCGGAGCGGATGCTTCGCGAGGA 432
Db      127 IleAlaYsArgValYsIleuThrProThrIleAlaGluArgThrValIleGlyGluGly 146
Qy      433 GACGCGACCATCTTCGCGGTGACGATACCGCC--ATCGGCGCGCTTCGCGCGCTCTGT 489
Db      147 AspGlySerAspIleuAlaValIleAspArgProAspIleGlyArgLeuGlyAlaMetCys 166
Qy      490 TGCTGGGAGCAATCCAGCCATTCGCAATACCGCATATACCGCGCGGAGCAAGGATC 549
Db      167 CysTrpGluHsIleuGlnProIleuSerIyTrAlaMetYrAlaGlnAsnGluGlnVal 186
Qy      550 CACGTGCGGTGCGCGGATTCAGCTCTATCGCGGATGCGCTATGCGATCGAGACG 609
Db      187 HisValAlaSerTrpProSerPheSerIleuYrAspProPheAlaHisAlaLeuGlyTrp 206
Qy      610 GAGGTCAATACCGCGCGGAGCAATCTACCGCGGTGAGGCGCGCTGTCTACGTCTGCG 669
Db      207 GluValAsnAsnAlaAlaSerArgValYrAlaValGluGlySerCysPheValIleuAla 226
Qy      670 TCGTGGGAGCAATCCAGCCATTCGCAATATTCGCAATATTCGCAATATTCGCAATATTC 729
Db      227 ProCysAlaTrpValSerGlnAlaMetIleAspArgIleuGlyAspArgProAspIySerSerIleu 246
Qy      730 ATGTTCTCAAGCGCGCGCGGTTCGCAATATTCGGGCGCGGAGCGCGCGCTGTG 789
Db      247 AlaLeuIleuHsAlaGlyGlyIleAlaValIleYrGlyProAspGlySerSerIleu 266
Qy      790 GCCGAGCGCGCTCCGCGAGACCGCAAGAGGATGCTGTCGCGGATATCTGAGCTCGAGATG 849
Db      267 AlaGluYsIleuArgProAspGlnGluGlyIleuYrAlaAspIleAspLeuGlyAla 286
```

QY 850 ATCGGTTGGCAAGCGCGCGCGGCGGCACTATTGACGGCCGCAAGTAACG 909  
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
Db 287 IleglyValAlaIalysaAmAlaIalAspProAlaGlyIstYserArgProAspValThr 306  
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
QY 910 CGGCTGCTGTGATTCGACGTCCGGCCCAAGCGCTC 945  
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
Db 307 ArgLeuLeuLeuAaenAlaenIleProIleTyrIleArgVal 318  
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
RESULT 13  
Q6RWF6\_92ZZZ PRELIMINARY; PRT; 336 AA.  
ID Q6RWF6\_92ZZZ  
AC Q6RWF6\_92ZZZ  
DT 05-JUL-2004 (TREMBlrel. 27, Created)  
DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)  
DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)  
DE Nitrilase (EC 3.5.5.7).  
GN ORFNames=BD5295;  
OS unclutered organism.  
OC unclassified; environmental samples.  
CX NCBI\_TaxID=155900;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RX PubMed=15066841; DOI=10.1128/AEM.70.4.2429-2436.2004;  
RA Robertson D.E., Chaplin J.A., Desantis G., Podar M., Menden M.,  
RA Chi E., Richardson T., Milan A., Miller M., Weiner D.P., Wong K.,  
RA McQuaid J., Farwell B., Preston L.A., Tan X., Sneed M.A., Keller M.,  
RA Mathur E., Kretz P.L., Burk M.J., Short J.M.;  
RT "Exploring nitrilase sequence space for enantioselective catalysis.";  
RL Appl. Environ. Microbiol. 70:2429-2436(2004).  
DR EMBL: AY487546; AAR97493.1; -; Genomic DNA.  
DR GO: 0018762; F:aliphatic nitrilase activity; IEA.  
DR GO: 0018810; F:hydrolase activity; acting on carbon-nitrog. . .; IEA.  
DR GO: 0006807; P:nitrogen compound metabolism; IEA.  
DR InterPro: IPR00132; N/CN\_hydrolase.  
DR Pfam: PF00795; CN\_hydrolase; 1.  
DR PROSITE: PS50263; CN\_HYDROLASE; 1.  
DR PROSITE: PS00920; NITRIL\_CHT\_1; 1.  
KW Hydrolase.  
SQ SEQUENCE 336 AA; 36221 MW; CC370C188F94CB72 CRC64;  
Alignment Scores:  
Pred. No.: 6.11e-56 Length: 336  
Score: 1070.00 Matches: 200  
Percent Similarity: 76.2% Conservative: 53  
Best Local Similarity: 60.2% Mismatches: 75  
Query Match: 56.9% Indels: 4  
DB: 2 Gaps: 2  
US-09-751-299-1 (1-1041) x Q6RWF6\_92ZZZ (1-336)  
QY 1 ATGTCGAGAGCATGATGATGCGCGCGCGGCGGCGAGCGCGCGGATGTCCTC 60  
::: ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
Db 3 IleserHisPro-----LysPheYalAlaIalValAlaGlnAlaGlyProAlaPheLeu 20  
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
QY 61 GATTCGACCGCACAGTCGAGAAAGCATGCGCGCTGATGACGAGCGCGCCACGACGAGAC 120  
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
Db 21 AspLeuAapGlyGlyValAlaGlnAlaValaIserLeuIleGlyGlnAlaIalAlaGlnGly 40  
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
QY 121 GTGGCGCGGATGCGATTCGCCAGACTTGATTCGCCGCTATTCCTTTGGATATGCTG 180  
::: ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
Db 41 AlaglnLeuIleAlaPheProGlnThrTrpIleProGlyIleProTrpHisThrTrpLeu 60  
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
QY 181 GGGCGCGCGCTGGGCGCATGCGC-----TTTCGTCAGCGCTATTTTCGAGATTCGCTC 234  
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
Db 61 GlyserProAlaIleAlaMetGlnIleGlyPheValAlaGlnAlaGlyIlePheAspAlaLeu 80  
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
QY 235 GTGGCGCGGACGACGATGCGAGCGCGCTGCGCGATGCGCGCGCGCCGACGCGCATG 294  
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
Db 81 ArgHisGlyserProGlnAlaGlnIleGlyIleSerGlyAlaAlaIalAlaGlnHisIleMet 100  
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
QY 295 GTTCGCGCGGCTATGAGGAGCGCGCGCGCGGCGCGCTCTATATGCGCGCAGCGCATCTTC 354

Db 101 ValSerLeuIleYheAlaGlnAlaArgAspGlyGlyThrLeuTyrIleAlaGlnTrpLeuIle 120  
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
QY 355 GGGCGCGGATGCGATTCGATGCCCGCGCGCGCAAGCTCAAGCTTACCATTGCGGAGCGC 414  
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
Db 121 GlyProAapIleGlnThrIleSerArgArgIleLeuYserProThrHisValAlaGlnArg 140  
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
QY 415 ACGGTGTGGCGAGGAGAGCGGACGACCATCTGCGGTGACGATACCGCATCGGCGCGC 474  
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
Db 141 ThrValPheGlyGlnGlyAspGlySerAspLeuSerValHisAspHisAlaLeuGlyArg 160  
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
QY 475 CTCGCGCGCTCTGTGCTGGGAGCAGCATCCAGCATTTGCGAATATGCGCATGTCGCC 534  
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
Db 161 IlegIleSerLeuYscYstrPglHisIleGlnProLeuSerIleYstrYalAlaMetYalAla 180  
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
QY 535 GCGGACGACAGAGTCCAGCTGCGGTGCGCGCGCGCATCTTACGCTTATCGCGGACGCC 594  
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
Db 181 GlnAsnGlnIleIleHisIleGlyAlaIleProSerPheSerLeuTyrGlnProPheAla 200  
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
QY 595 TATGCGCTCGGACCGGAGTCATACCGCGCAAGCATCTTACCGGTCGAGGCGCGC 654  
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
Db 201 AsnAlaLeuSerProGlnValaAsnIleAlaValaSerArgValIleYalAlaGlnGlyGln 220  
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
QY 655 TGCTACGCTGCGCGCTGCGCGGCGGCGCGGCGGCGGAGTATGATCAAGTATTCGTCGAT 714  
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
Db 221 CysPhePheLeuAlaProCysAlaIleThrValaSerAspAlaMetIleGlnThrLeuCysAsp 240  
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
QY 715 ACGCGCGACAGAGATGCTCTCAGAGCGCGCGCGGCTTTGCGATGATTTTCGGGCGC 774  
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
Db 241 ThrProGlnGlySerGlnGlyLeuIleAlaGlnGlyGlyGlnHisAlaIlePheGlyPro 260  
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
QY 775 GAGCGCGCGCGCTGCGCGGCGCGCGCGCGCGCGGACGAGAGGAGCTGTCGCGCAT 834  
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
Db 261 AspGlySerLeuLeuThrProThrValaAlaAspHisThrGlnGlyLeuLeuTyrAlaGln 280  
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
QY 835 ATGACCTTCGATGATGCGGCTTGCGCGCAAGCGCGCGCGCATGCGCGCGCACTATTC 894  
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
Db 281 LeuAspLeuGlyAlaIleSerIleAlaIleYserAlaAlaAspProAlaGlyHisIleYser 300  
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
QY 895 CGGCGCGACGTAAGCGCGCTGCTGCGATGACGTCGCGCGCCACGCGTCGTCACGCTT 954  
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
Db 301 ArgProAapValThrArgLeuLeuLeuAsnGlnThrProSerIleYargValGlnAsnMet 320  
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
QY 955 GATCGCGCATTCGACCGCGCAAGCGAGACGCGC 990  
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
Db 321 ValLeuProLeuGlnThrValThrGlnProGlnGly 332  
||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :  
RESULT 14  
Q6RWF7\_92ZZZ PRELIMINARY; PRT; 335 AA.  
ID Q6RWF7\_92ZZZ  
AC Q6RWF7\_92ZZZ  
DT 05-JUL-2004 (TREMBlrel. 27, Created)  
DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)  
DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)  
DE Nitrilase (EC 3.5.5.7).  
GN ORFNames=BD5326;  
OS unclutered organism.  
OC unclassified; environmental samples.  
CX NCBI\_TaxID=155900;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RX PubMed=15066841; DOI=10.1128/AEM.70.4.2429-2436.2004;  
RA Robertson D.E., Chaplin J.A., Desantis G., Podar M., Menden M.,  
RA Chi E., Richardson T., Milan A., Miller M., Weiner D.P., Wong K.,  
RA McQuaid J., Farwell B., Preston L.A., Tan X., Sneed M.A., Keller M.,  
RA Mathur E., Kretz P.L., Burk M.J., Short J.M.;  
RT "Exploring nitrilase sequence space for enantioselective catalysis.";  
RL Appl. Environ. Microbiol. 70:2429-2436(2004).  
DR EMBL: AY487545; AAR97492.1; -; Genomic DNA.  
DR GO: 0018762; F:aliphatic nitrilase activity; IEA.  
DR GO: 0018810; F:hydrolase activity; acting on carbon-nitrog. . .; IEA.  
DR GO: 0006807; P:nitrogen compound metabolism; IEA.  
DR InterPro: IPR00132; N/CN\_hydrolase.



OY	262	CTGGGCGAATGCAGCGCGCGCGCGCGCAAGCATGTGTCGGCGCGCGTAAATGACGCGCGCG	321
Db	87	LeuAlaGluAlaAlaArgAlaSerAlaIleThrValValThrCylAspGlyValLysAlaGluArg	106
OY	322	GCGCGCAAGCCTCTATATGGCGCGCGCAAGCGATCTTCGAGCGCGCGATGGCCATGATCGCGCGG	381
Db	107	GlyGlySerLeuThrMetGlyMetGlyAlaIleIleGlyAlaAspGlyGlyValLysAlaAla	126
OY	382	CGCCCGCAAGCTCAAGCGTACCATGCGCGAGCGCGACCGTGCTGGCGCGAGGACGCGCGACG	441
Db	127	ArgAlaGlyLeuLeuLysProThrHisValGlyAlaGlyThrValPheGlyLysSerArgGlySer	146
OY	442	CATTCGCGGCTGCACATATCGCGCGCATCGGAGCGCGCTGAGCGCGCGCTGTTGGTCCGAAGAAC	501
Db	147	AsnLeuThrValValAlaSerThrGlyMetGlyAlaGlyLeuGlyAlaLeuGlySerCysThrGlnHis	166
OY	502	ATCCAGCAACTGTGCAATAATACGSCATGTAACGCGCGCGCGCGCAAGACAGGTCACAGCTCGCGTGS	561
Db	167	LeuGlnProLeuThrLysThrAlaMetCysTrpSerGlnHisGluGlnIleHisValAlaAlaAla	186
OY	562	TGGCGGAGCTTACAAGCCTCTATGCGCGCGCACAGCGCTATGCGCTGCGACCGGAGCGCAATACCS	621
Db	187	TrpProSerPheSerValTrpArgGlyAlaAlaAlaGlyAlaLeuGlyProGluValAlaSerThr	206
OY	622	GCCGCGAAGCGAATGATCAAGCGCGTCCAGCGCGCGCGCGCTAGCGTGGCGCGCGTCCGCGCGACG	681
Db	207	GlyAlaAlaAlaArgGlnTrpAlaValGlyGlyGlnCysPheValLysSerProCysAlaValAl	226
OY	682	GTTTTCGCGGAGATGATCAAGGATATTGCTGATATACGCGCGCAAGACAGAGATGTTCTCCAGAG	741
Db	227	IleAspGlnAlaGlyValGlyLeuPheCysValAspThrProAlaLysArgGlyLeuLeuLeu	246
OY	742	GCGCGCGCGCGTTTGGCCATGATTTTGGCGCGCGCGCGACGCGCGCGCGCTGCGCGACCGCGTC	801
Db	247	ProGlyGlyGlyPheAlaGlnIleLeuTrpGlyProAspGlyArgGlyLeuGlnIyThrAlaLeu	266
OY	802	CCGAGAGACCGAAGAGAGGATGCTGGCGCGCGCGCATATGACCTGGCATATGATCCGCTGGCGG	861
Db	267	ProGluThrGlyGlyGlyLeuValTrpAlaSerLeuGlnAlaSerAlaValAlaValAlaAla	286
OY	862	AAGCGCGCGCGCGATCGCGCGCGCGCACTTATCAAGCGCGCGCGAGTAGACGCGCGCTGCTGTG	921
Db	287	LysSerAlaAlaAspProValGlyHisValTrpSerArgProAspValLeuGlnIleuLeuThr	306
OY	922	GATGACAGCTCGGCGCCAA-----CGCGTCGTCACGCTTGATGCGCGCATTCGACCGCGCA	975
Db	307	AspProAlaGlyProArgSerValValArgGlnValAlaAlaLeuSerValAlaSerProAlaGln	326
OY	976	AACGAGACAGAGGCGGCGCGCGCGCGCGCGCGC	1008
Db	327	SerAlaAspArgAlaGln---ProAlaValAlaG	336

Search completed: April 27, 2006, 01:35:33  
Job time : 263.071 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - protein search, using frame\_n2p model

Run on: April 27, 2006, 01:06:26 ; Search time 34.8013 Seconds  
(without alignments)  
2628.598 Million cell updates/sec

Title: US-09-751-299-1

Perfect score: 1881  
Sequence: 1 atgcgcgagccatgcagca.....gcccgcgcgcgcgcagtag 1041

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 4886326

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODE=frame+np.model -DEV=xlp  
-Q=/abs/ABSSWEB.spool/US09751299/runat\_26042006\_090032\_18619/app\_query.fasta\_1  
-DB=A.Geneseq -OPMT=faetan -SUFFIX=rag -MINMATCH=0.1 -LOOPCL=0 -LOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blomsum62 -TRANS=human40.cdi -LIST=45  
-DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=ptc -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000 -HOST=abs802p  
-USER=US09751299\_@CGN\_1\_1\_476\_@runat\_26042006\_090032\_18619 -NCPU=6 -ICPU=3  
-NO\_MMAB -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG -DEV TIMEOUT=120  
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

A\_Geneseq\_21:\*  
1: geneseqp1980s:\*  
2: geneseqp1990s:\*  
3: geneseqp2000s:\*  
4: geneseqp2001s:\*  
5: geneseqp2002s:\*  
6: geneseqp2003as:\*  
7: geneseqp2003bs:\*  
8: geneseqp2004s:\*  
9: geneseqp2005s:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length DB	ID	Description
1	1806	96.0	346	4	AAE05488 Nitriase
2	1806	96.0	346	7	ADc24117 Protein s
3	1806	96.0	346	8	ADh36218 Chemical
4	1806	96.0	346	8	ADg93917 Nitriase
5	1806	96.0	346	8	ADi62516 Nitriase
6	1806	96.0	346	8	ADi64637 Nitriase
7	1464	77.8	341	7	ADc23767 Protein s
8	1464	77.8	341	8	ADh35868 Chemical
9	1464	77.8	341	8	ADg93569 Nitriase

10	1464	77.8	341	8	ADi62166 Nitriase
11	1464	77.8	341	8	ADi64287 Nitriase
12	1209	64.3	348	7	ADc23749 Protein s
13	1209	64.3	348	8	ADh35850 Chemical
14	1209	64.3	348	8	ADg93551 Nitriase
15	1209	64.3	348	8	ADi62148 Nitriase
16	1209	64.3	348	8	ADi64269 Nitriase
17	1165	61.9	353	7	ADc23945 Protein s
18	1165	61.9	353	8	ADh36046 Chemical
19	1165	61.9	353	8	ADg93747 Nitriase
20	1165	61.9	353	8	ADi62344 Nitriase
21	1165	61.9	353	8	ADi64465 Nitriase
22	1155	61.4	353	7	ADc23965 Protein s
23	1155	61.4	353	8	ADh36066 Chemical
24	1155	61.4	353	8	ADg93767 Nitriase
25	1155	61.4	353	8	ADi62364 Nitriase
26	1155	61.4	353	8	ADi64485 Nitriase
27	1148	61.0	333	7	ADc24083 Protein s
28	1148	61.0	333	8	ADh36184 Chemical
29	1148	61.0	333	8	ADg93883 Nitriase
30	1148	61.0	333	8	ADi62482 Nitriase
31	1148	61.0	333	8	ADi64603 Nitriase
32	1132.5	60.2	358	8	ADc23903 Protein s
33	1132.5	60.2	358	8	ADh36004 Chemical
34	1132.5	60.2	358	8	ADg93705 Nitriase
35	1132.5	60.2	358	8	ADi62302 Nitriase
36	1132.5	60.2	358	8	ADi64423 Nitriase
37	1130	60.1	354	7	ADc23835 Protein s
38	1130	60.1	354	8	ADh35936 Chemical
39	1130	60.1	354	8	ADg93637 Nitriase
40	1130	60.1	354	8	ADi62234 Nitriase
41	1130	60.1	354	8	ADi64355 Nitriase
42	1125	59.8	345	7	ADc24067 Protein s
43	1125	59.8	345	8	ADh36168 Chemical
44	1125	59.8	345	8	ADg93868 Nitriase
45	1125	59.8	345	8	ADi62466 Nitriase

#### ALIGNMENTS

RESULT 1	AAE05488
ID	AAE05488 standard; protein; 346 AA.
XX	
AC	AAE05488;
XX	
DT	25-SEP-2001 (first entry)
XX	
DE	Nitriase protein #1.
XX	
KW	Nitriase; enantiomer; alpha-substituted carboxylic acid;
KW	stereoselective; hydrolysis; amino nitriase; cyanohydrin.
XX	
OS	Unidentified.
XX	
PN	WO200148175-A2.
XX	
PD	05-JUL-2001.
XX	
PF	29-DEC-2000; 2000WO-US035555.
XX	
PR	29-DEC-1999; 99US-0173609P.
XX	
PR	07-DEC-2000; 2000US-0254414P.
XX	
PA	(DIVE-) DIVERSA CORP.
XX	
PI	(MADD-) MADDEN D.
XX	
DR	Madden M, Weiner DP, Chaplin JA;
XX	
DR	WPI; 2001-465211/50.
XX	
PT	N-PSDB; AAD11166.
	Producing alpha-substituted carboxylic acid enantiomers by contacting

PT aldehyde or ketone with cyanide, ammonia compound or its salt or amine  
PT and hydrolyzing the resulting amino nitrile or cyanohydrin with  
PT Nitrilase.

PS Claim 29; Page 84-85; 87pp; English.

XX The present invention relates to methods for producing an  
CC enantiomerically pure alpha-substituted carboxylic acid. The method  
CC involves contacting an aldehyde or ketone with a cyanide containing  
CC compound and an ammonia-containing compound or an ammonium salt or amine,  
CC and stereoselectively hydrolyzing the resulting amino nitrile or  
CC cyanohydrin intermediate with a nitrilase or a polypeptide having  
CC nitrilase activity. The present sequence is a nitrilase protein #1

XX Sequence 346 AA;

Alignment Scores:

Pred. No.:	1,53e-148	Length:	346
Score:	1806.00	Matches:	346
Percent Similarity:	100.0%	Conservative:	0
Best Local Similarity:	100.0%	Mismatches:	0
Query Match:	96.0%	Indels:	0
DB:	4	Gaps:	0

US-09-751-299-1 (1-1041) x ABE05488 (1-346)

QY 1 ATGTGGAGCCCGTGAACGAATATCGGGCGCGGCGGTGAGCGCCGCGGTCTTC 60  
DB 1 MetSerGluProMetThrLysTyrArgGlyAlaAlaValGlnAlaIleProValPheLeu 20  
QY 61 GATCTGACCGCGCAGTCGAGAAAGCATCGGCGCTGATGAGAGCGCGCAAGCAGAC 120  
DB 21 AspleuAspArgThrValGlnLysAlaIleGlyLeuIleGlnAlaAlaLysGlnAsp 40  
QY 121 GTGCGCTGATGCGATTCCTCCAGACTTGATTCGCGCTATCCCTTTGGATATGCTG 180  
DB 41 ValArgLeuIleAlaPheProGluThrTrrPleProGlyTyrProPheTrpIleTrpLeu 60  
QY 181 GGGCGCCCGGCTTGGGCGATGCGCTTCGTCAGCGCTATTTGGAATTCGCTCGCGCG 240  
DB 61 GlyAlaProAlaTrpGlyMetArgPheValGlnArgTyrPheGlnLysSerLeuValArg 80  
QY 241 GCGAGCAGCAGTGGAGCGCCCTGCGGATGCGCGCGCGCGCGCGCATGTCATGCTG 300  
DB 81 GlySerLysGlnTrpGlnIleAlaLeuAlaAspAlaAlaArgArgHisGlyMetHisValVal 100  
QY 301 GCGGCTATATGCGAGCGCGCGCGCGCGCGCTTATATGCGCGCGCATCTTGGCGCC 360  
DB 101 AlaGlyTyrSerGluTrpGlnArgGlySerLeuTyrMetGlyGlnAlaIlePheGlyPro 120  
QY 361 GATGCGCATGTCATCGCGCGCGCGCGCGCATCAAGCTCAAGCTTACCATGCGAGCGCACCGTG 420  
DB 121 AspGlyAspLeuIleAlaAlaArgGlyLysLeuLysProThrHisAlaGlnArgThrVal 140  
QY 421 TTCGCGAGGAGAGCGAGCGCATCTTCGCGGTGACAGTACCGCATCGCGCGCTCGGC 480  
DB 141 PheGlyGlyGlyAspGlySerHisLeuAlaValHisAspThrAlaIleGlyArgLeuGly 160  
QY 481 GCGCTGCTGTCGCGGAGCAGCATCCGCGCATTCGAAATACCGCATGACCGCGCGCAGC 540  
DB 161 AlaLeuCySerTrpGlnHisGlnPheLeuSerLysTyrTrpAlaLeuTyrAlaAlaAsp 180  
QY 541 GAACAGGTCACAGCTCGTCGTCGCGAGCTTACGCTTATGCGCGCATGCGCTATGCG 600  
DB 181 GlnGlnValHisValAlaSerTrpProSerPheSerLeuTyrArgGlyMetAlaTyrAla 200  
QY 601 CTGGAACCGGAGTCAATACCGCGCAAGCAGATCTACGCGGTGCAAGCGCGCTCTAC 660  
DB 201 LeuGlyProGlnValAsnThrAlaAlaSerGlnIleTyrAlaValGlnGlyCysTyr 220  
QY 661 GTGCTGCGCGTGGCGCGCGCGCTTTCGCGGAGATGATCAAGTATGATGATGATGATG 720  
DB 221 ValLeuAlaSerCysAlaThrValSerProGlnMetIleLysValIleuValAspTrpPro 240

QY 721 GACAAAGATGTTCTCTCAAGCGCGCGCGGCTTTGTCATGATTTTGGCGCCGACGCG 780  
DB 241 AspLysGluMetPheLeuIleAlaGlyGlyPheAlaMetIlePheGlyProAspGly 260  
QY 781 CGCGCCCTTGGCGCGCGCGCTCCCGGAGCCGAAAGGAGCATGCTGTGTGCGCGATTCAC 840  
DB 261 ArgAlaLeuAlaGlnProLeuProGlnTrpGlnGlyLeuLeuValAlaAspIleAsp 280  
QY 841 CTGCGCATGATGCGCTTGGCGCAAGCGCGCGCGCGCATCCGCGCGCGCATATTCACGCGCC 900  
DB 281 LeuGlyMetIleAlaLeuAlaLysAlaAlaAspProAlaGlyHisTyrSerArgPro 300  
QY 901 GACGTAACGCGCGCTGCTGCTGATGACGTCGCGCGCAAGCGCTGTCACGCTTATGCC 960  
DB 301 AspValThrArgLeuLeuLeuAspArgProAlaGlnArgValAlaThrLeuAspAla 320  
QY 961 GCATTCGAACCGCAAAAGAGGAGCAAGGCGAGCGCGCGCTGCGCGGTGGCGGAA 1020  
DB 321 AlaPheGluProGlnAsnGlnLysGlyAspAlaProAlaLeuArgValAlaGln 340  
QY 1021 AGCGCGCGCGCGCGCGCAG 1038  
DB 341 SerAlaAlaAlaGln 346

RESULT 2  
ADC24117  
ID ADC24117 standard; protein; 346 AA.  
XX  
AC ADC24117;  
XX  
DT 18-DEC-2003 (first entry)  
XX  
DE Protein sequence (SeqID 384) exhibiting nitrilase activity.  
XX  
KW enzyme; nitrilase; nitrile; cyanohydrin; ammonia; biocatalyst;  
XX nitrilomer; chiral medicine.  
XX  
OS Unidentified.  
XX  
PN WO2003000840-A2.  
XX  
PD 03-JAN-2003.  
XX  
PF 15-MAY-2002; 2002MO-US015983.  
XX  
PR 21-JUN-2001; 2001US-0300189P.  
XX 30-JUL-2001; 2001US-030906P.  
XX 22-JAN-2002; 2002US-0351336P.  
XX  
PA (DIVE-) DIVERSA CORP.  
PA (MADD/) MADDEN D.  
XX  
PI Madden M, Desantis G, Chaplin JA, Weiner DP, Milan A, Chi E;  
PI Short JM, Burk M;  
XX  
DR MPI: 2003-201417/19.  
XX N-PSDB; ADC24116.  
XX  
PT Novel nitrilase polypeptide, useful for making (R)- or (S)-ethyl-4-cyano-  
PT 3-hydroxybutyric acid or (R)- or (S)-mandelic acid or (S)- or (R)-phenyl  
PT lactic acid derivative and for producing pharmaceutical composition, and  
PT food additive.  
XX  
PS Claim 40; SEQ ID NO 384; 560pp; English.

XX The present invention relates to nitrilases and the nucleic acids that encode  
XX these enzymes thereof. Specifically, it refers to polypeptides that  
XX exhibit nitrilase activity, i.e. the ability to directly hydrolyse  
XX nitriles or cyanohydrins into their corresponding carboxylic acids and  
XX ammonia. Nitrilases have commercial utility as biocatalysts for use in  
XX the synthesis of enantiomerically pure aromatic and aliphatic amino  
XX acids, as well as hydroxy acids, which are important for the development



Best Local Similarity: 100.0%  
 Query Match: 96.0%  
 DB: 8  
 Gaps: 0

US-09-751-299-1 (1-1041) x ADH36218 (1-346)

```

QY 1 ATGTCGAGAGCCCATGACGAGATGCGCGCGCGGTCGACGCGCGCGGTCGTC
DB 1 MetSerGluProMetThrLysTrpGlyAlaValGlnAlaAlaProValPheLeu
QY 61 GATTCGACCGCACAGTCGAGAAAGCGATCGGCTGATCGACGCGCGCCAGCAGAC
DB 21 AspLeuAparGThrValGluValAlaIleGlyLeuIleGlnAlaIleValGlnAsp
QY 121 GTGGCCCTGATCGATTCGCCAGACTTGATTCGCCGCTATCCCTTTGGATATGCGTG
DB 41 ValArgLeuIleAlaPheProGluThrTriPleProGlyTrpPropheTriPleTrpLeu
QY 181 GCGCGCGCGGCTTGGGCGATGCGCTTGCCTCAGCGGCTATTTCGAAATTCGCTGTCGCG
DB 61 GlyAlaProAlaTrpGlyMetArgPheValGlnArgTrpPheGlnAbaSerLeuValArg
QY 241 GCGCAGCAAGAGTCGACAGGCCCTTGGCGAGTCGCGCGCGCCAGCGCATGATGTCGTG
DB 81 GlySerLysGlnTrpGlnAlaLeuAlaAspAlaAlaArgArgHisGlyMetHisValVal
QY 301 GCCCGCTATAGCGCGCGCGCGCGCGCGCTCTATATGCGCGCGCGAGATTCGCGCGCC
DB 101 AlaGlyTrpSerGluArgAlaGlyGlySerLeuTrpMetGlyGlnAlaIlePheGlyPro
QY 361 GATGGCGATCTGATCGCGCGCGCGCGCGAGCTCAAGCTTCAAGCTTCCATGCGGACCGCGTG
DB 121 AspGlyAspLeuIleAlaAlaArgArgLysLeuSerThrHisAlaGlnAbaGlnThrVal
QY 421 TTGCGCGCAGGAGACCGCGCAGCCATCTCGCGGTGACGATACCGCGCATCGGCGCTTCGCG
DB 141 PheGlyGlnGlyAspGlySerHisLeuAlaValHisAspThrAlaIleGlyArgLeuGly
QY 481 GCGCTCTGTTGTCGAGCAGCATCCAGCATTTGTCGAATACCGCATGTCGCGCGCGGAC
DB 161 AlaLeuLysCysTrpGlnHisIleGlnProLeuSerLysTrpAlaMetCysTrpAlaAlaAsp
QY 541 GAACAGGTCCACGTCGCGGTGCGCGCGAGCTTACGCTTATCGCGCATGAGCTTATCGC
DB 181 GlnGlnValHisValAlaSerTrpProSerPheSerLeuTrpArgGlyMetAlaTrpAla
QY 601 CTGCGACCGGAGGTCATACCGCGCGCAGGACGATCTACCGCGGTTCGAGCGCGCTGTAC
DB 201 LeuGlyProGlnValAsnThrAlaAlaSerGlnIleTrpAlaValGlnGlyCysTrp
QY 661 GTGCTGCGCTGTCGCGCAGCCGTTTCGCGCGAGATGATCAAGGTAATTGGATACGCGCC
DB 221 ValLeuAlaSerCysAlaThrValSerProGluMetCysValLeuValAspThrPro
QY 721 GACACGAGATGTTCTCAAGCGCGCGCGGCTTTGCGATGATTTTCGCGCGCAGCGC
DB 241 AspLysGluMetCysPheLeuLysAlaGlyGlyGlyPheAlaMetCysIlePheGlyProAspGly
QY 781 CCGCGCCCTGGCGCAGCGCGCTCCGCGGACCGAGAGAGGATGTCGTGTCGCCGATATCGAC
DB 261 ArgAlaLeuAlaLeuProLeuProGlnThrGlnGlnGlyLeuLeuValAlaAspIleAsp
QY 841 CTCGCGATGATCGCGTTGGCGCAAGCGCGCGCGCGATCCGCGCGCGCATATTCGCGCGC
DB 281 LeuGlyMetCysIleAlaLeuAlaLysAlaAlaAlaAspProAlaGlyHisTrpSerAlaArgPro
QY 901 GAGCTAACCGCGGTGCTGCTGATTCACGTCGCCCGCAACGCGTGTCAAGCTTGATGCC
DB 301 AspValThrArgLeuLeuLeuAparGTrpAlaGlnArgValAlaThrLeuAparAla
QY 961 GCATTCGAACCGCAAAACGAGACAGAGGCGGACGCGCGCGCGCTGCGCGGTGGGCGGAA
DB 321 AlaPheGluProGlnAsnGlnAspLysGlyAspAlaProAlaLeuArgValAlaAlaGln
  
```

QY 1021 AGCGCGCGCGCGCGCAG 1038  
 DB 341 SerAlaAlaAlaAlaGln 346

#### RESULT 4

ADG93917  
 ID ADG93917 standard; protein; 346 AA.

AC ADG93917;  
 XX

DT 11-MAR-2004 (first entry)  
 XX

DE NitriIase enzyme amino acid sequence SegID384.  
 XX

KM nitriIase; nitrite; carboxylic acid; chemical process; pH; temperature;  
 XX enantioselective transformation; enzyme.

OS Unidentified.  
 XX

PN WO2003097810-A2.  
 XX

PD 27-NOV-2003.  
 XX

PF 15-MAY-2003; 2003WO-US015712.  
 XX

PR 15-MAY-2002; 2002US-00146772.  
 XX

PR 09-SEP-2002; 2002US-00241742.  
 XX

PA (DIVE-) DIVERSA CORP.  
 XX

PI Desantis G, Short JM, Burk M, Wong K, Farwell R, Chatman K;  
 XX

DR WPI; 2004-090637/09.  
 XX

DR N-PSDB; ADG93916.  
 XX

PT New isolated or recombinant nucleic acid encoding a polypeptide having  
 XX nitriIase activity, useful for screening enantioselective transformation.

PS Claim 44; SEQ ID NO 384; 295bp; English.  
 XX

CC This invention is related to a novel isolated or recombinant nucleic acid  
 CC encoding a protein having nitriIase activity. NitriIase's are capable of  
 CC converting nitrite's directly to carboxylic acids and have great  
 CC potential for use in industrial chemical processes. The isolated  
 CC nitriIase proteins of the invention have increased activity and stability  
 CC at increased pH and temperature when compared to those conventionally  
 CC used. In addition, the nucleic acid of the invention is useful for  
 CC screening enantioselective transformation. The present sequence is that  
 CC of a nitriIase enzyme of the invention.

SQ Sequence 346 AA;

#### Alignment Scores:

Pred. No.:	1,53e-148	Length:	346
Score:	1806.00	Matches:	346
Percent Similarity:	100.0%	Conservative:	0
Best Local Similarity:	100.0%	Mismatches:	0
Query Match:	96.0%	Indels:	0
DB:	8	Gaps:	0

US-09-751-299-1 (1-1041) x ADG93917 (1-346)

```

QY 1 ATGTCGAGAGCCCATGACGAGATGCGCGCGCGGTCGACGCGCGCGGTCGTC
DB 1 MetSerGluProMetThrLysTrpGlyAlaValGlnAlaAlaProValPheLeu
QY 61 GATTCGACCGCACAGTCGAGAAAGCGATCGGCTGATCGACGCGCGCCAGCAGAC
DB 21 AspLeuAparGThrValGluValAlaIleGlyLeuIleGlnAlaIleValGlnAsp
QY 121 GTGGCCCTGATCGATTCGCCAGACTTGATTCGCCGCTATCCCTTTGGATATGCGTG
  
```

Db ValArgLeuIleAlaPheProGluThrTrpIleProGlyTyrProPheTrpIleTrpLeu 60  
QY 181 GGGCGCGCGGTTGGGGCATCCGCTTGTCAGCGCTATTTCGAATTGCTGTGGCC 240  
Db 61 GlyAlaProAlaTrpGlyMetArgPheValGlnArgTyrPheGluAenSerLeuValArg 80  
QY 241 GGGAGCAAGCGATGGCGGCGCTGGCGGATGCGGCGCGCGCGCAAGGCAATGCTGTG 300  
Db 81 GlySerIysGlnTrpGlnAlaLeuAlaAspAlaAlaArgArgHisGlyMetHisVal 100  
QY 301 GCCCGCTATAGCGAGCGCGCGGCGAGCCTCTATATGGCGCGAGCATCTTGGCC 360  
Db 101 AlaGlyTyrSerGluArgAlaGlyGlySerLeuTyrMetGlyGlnAlaIlePheGlyPro 120  
QY 361 GATGGCATCTGATCGCGCGCGCGCGCAAGCTTCAAGCTTCAAGCGGAGCGCAAGCTG 420  
Db 121 AspGlyAspLeuIleAlaAlaArgArgLeuLysProThrHisAlaGluArgThrVal 140  
QY 421 TTGGCGAGGAGAGCGGAGCATCTCGCGGTGACATACCGGCATCGCGCGCGCTGGC 480  
Db 141 PheGlyGluGlyAspGlySerHisLeuAlaValHisAspThrAlaIleGlyArgLeuGly 160  
QY 481 GCGCTGTGTGCTGGAGCAATCCAGCATCTGTCTGAATAGCGCATGTAGCGCGCGAC 540  
Db 161 AlaLeuCysCysTrpGluHisIleGlnProLeuSerIlyTyrAlaMetTyrAlaAlaAsp 180  
QY 541 GAACAGGTCCAGTGGCGGTGGCGCGCAAGCTTCAAGCTTATCGCGGCGATGGCTATGCG 600  
Db 181 GluGlnValHisValAlaSerTrpProSerPheSerLeuTyrArgGlyMetAlaTyrAla 200  
QY 601 CTGGAGCGGAGGTCATATACCGCGCAAGCAAGCAATCAACGGGTCGAGCGGCGGTGTAC 660  
Db 201 LeuGlyProGluValaIleThrAlaAlaSerGlnIleTyrAlaValGluGlyCysTyr 220  
QY 661 GTGTGCGCGTGTGTCGCGGACCGTTTCCCGGAGATGATCAAGATTGATGATACCGCC 720  
Db 221 ValLeuAlaSerCysAlaIleThrValSerProGluMetIleLysValLeuValAspThrPro 240  
QY 721 GACAAAGATGTTCTCAAGCGCGCGCGCGGTTTTCGATGATTTTCGGGCGCGACGCG 780  
Db 241 AspArgGluMetPheLeuLysAlaGlyGlyGlyPheAlaMetIlePheGlyProAspGly 260  
QY 781 CGCGCGCTGGCGCGCGCGCTCCCGAGACCAAGAGGACCTGCTGCGCGCATATGAC 840  
Db 261 ArgAlaLeuAlaGluProLeuProGluThrGlnGluGlyLeuLeuValAlaAspIleAsp 280  
QY 841 CTGGCATGATCGCGTGGCGCAAGCGGCGCGCATCGCGGCGCGCACTATTTCAGCGCCC 900  
Db 281 LeuGlyMetIleAlaLeuAlaLysAlaAlaAlaAspProAlaGlyHisTyrSerArgPro 300  
QY 901 GACGTAAAGCGGCTGCTGTGATGACAGTCCGCGCGCAAGCGGCTGTCACGCTTGATGCC 960  
Db 301 AspAlaIleThrArgLeuLeuLeuAspArgArgProAlaGlnArgValValThrLeuAspAla 320  
QY 961 GCATTCCAAACCGCAAGCAAGGAGCAAGCGCGCGCGCTGCGCGCTGTGGCGGAA 1020  
Db 321 AlaPheGluProGlnAenGluAspLysGlyAspAlaProAlaLeuArgValValAlaGlu 340  
QY 1021 AGCGCGCGCGCGCGCAG 1038  
Db 341 SerAlaAlaAlaAlaGln 346

KM (R)-ethyl 4-cyano-3-hydroxybutyric acid; epichlorohydrin;  
KM 3-hydroxyglutaronitrile; 3-hydroxyglutaronitrile;  
KM 4-cyano-3-hydroxybutyric acid; ethyl-4-cyano-3-hydroxybutyric acid;  
KM mixed hyperlipidaemia; homozygous familial hypercholesterolaemia;  
KM antilipemic; enzyme.  
OS Unidentified.  
XX  
XX W02003106415-A2.  
XX  
XX  
XX 24-DEC-2003.  
XX  
XX 13-JUN-2003; 2003MO-US018840.  
XX  
XX 13-JUN-2002; 2002US-0389317P.  
XX 28-JUN-2002; 2002US-0392944P.  
XX  
XX (DIVE-) DIVERSA CORP.  
PI Burk M, Desantis G, Morgan B, Zhu Z;  
XX  
XX WPI; 2004-090821/09.  
DR N-PSDB; ADI62515.  
XX  
XX Preparation of atorvastatin comprises catalytic conversion of 3-  
PT hydroxyglutaronitrile by polypeptide with nitrilase activity, converting  
PT obtained 4-cyano-3-hydroxybutyric acid to ethyl-4-cyano-3-hydroxybutyric  
PT acid and forming atorvastatin.  
PS Claim 41; SEQ ID NO 384; 253bp; English.  
XX  
XX The present invention relates to a method for preparing an atorvastatin  
CC intermediate known as (R)-ethyl 4-cyano-3-hydroxybutyrate ((R)-ethyl 4-  
CC cyano-3-hydroxybutyric acid). The method comprises optionally converting  
CC epichlorohydrin or equivalent to 3-hydroxyglutaronitrile, catalytic  
CC conversion of 3-hydroxyglutaronitrile or equivalent to 4-cyano-3-  
CC hydroxybutyric acid with a polypeptide having nitrilase activity,  
CC converting 4-cyano-3-hydroxybutyric acid to ethyl-4-cyano-3-hydroxybutyric  
CC acid, and converting this to (R)-ethyl 4-cyano-3-hydroxybutyrate. The  
CC method involves whole cell processes, cell lysate process, "one pot"  
CC processes, and "multi-pot" processes using a variety of parameters.  
CC Atorvastatin is used, in conjunction with dietary restriction, in the  
CC management of hyperlipidaemia, including hypercholesterolaemia, mixed  
CC dyslipidaemia and homozygous familial hypercholesterolaemia. The present  
CC sequence represents a nitrilase polypeptide obtained from an  
CC environmental sample.  
XX  
XX  
SQ Sequence 346 AA:  
XX  
XX  
XX Alignment Scores:  
Pred. No.: 1,53e-148 Length: 346  
Score: 1806.00 Matches: 346  
Percent Similarity: 100.0% Conservative: 0  
Best Local Similarity: 100.0% Mismatches: 0  
Query Match: 96.0% Indels: 0  
DB: 8 Gaps: 0  
US-09-751-299-1 (1-1041) x ADI62516 (1-346)  
QY 1 ATGTGGAGCCCATGACGAAGTATCGCGGCGCGGATGACGCGCGCGGTTCTTC 60  
Db 1 MetSerGluProMetThrIlystyTArgGlyAlaAlaValaGlnAlaAlaProValPheLeu 20  
QY 61 GATTCGACCGCACAGTCGAGAAAGCATCGGCTGATCGAGCGCGCGCAAGCAGAC 120  
Db 21 AspLeuAspArgThrValGluLysAlaIleGlyLeuIleGluGlnAlaLysGlnAsp 40  
QY 121 GTGGCGCTGATGCGATTCGACGAGACTTGATTCGCGGCTATCCCTTTGGATATGCTG 180  
Db 41 ValArgLeuIleAlaPheProGluThrTrpIleProGlyTyrProPheTrpIleTrpLeu 60  
QY 181 GGGCGCGCGCTTGGGCGATGCGCTTGTTCAGCGCATTTGAGAAATTCGCTGTGCGC 240



Db 41 ValArgLeuIleAlaPheProGluThrTriPleProGlyTrpProPheTriPleTrpLeu 60  
QY 181 GGGCGCGCGCTTGGGGCATGCGCTTGGTCCAGCGCTATTTCGAAATTCGCTCGTGGC 240  
Db 61 GlyAlaProAlaTrpGlyMetArgPheValGlnArgTrpPheGlnAseSerLeuValArg 80  
QY 241 GGGCAGCAGGAGTGGCGGCGCGCTTGGGGCATGCGCTTGGTCCAGCGCTATTTCGAAATTCGCTCGTGGC 300  
Db 81 GlySerIysGlnTrpGlnAlaLeuAlaSerPalaAlaArgArgHisGlyMetHisValVal 100  
QY 301 GCCGGCTATAGCGAGCGCGCGCGCGAGCTTATATAGCGCGCGCGCTTGGGGCATGCGCTCGTGGC 360  
Db 101 AlaeIlyTrserGlnArgAlaGlyGlySerLeuTrpMetGlyGlnAlaIlePheGlyPro 120  
QY 361 GATGGCGATTCGATCGCGCGCGCGCGAGCTCAAGCTTACCGCGAGCGCGAGCGCGCTG 420  
Db 121 AspGlyAspLeuIleAlaAlaArgArgIlySerLeuYsProThrHisAlaGlnArgThrVal 140  
QY 421 TTGGCGAGGAGAGCGCGAGCGCGATCTGCGGTGACATTCGCGCGCGCGCTGCGG 480  
Db 141 PheGlyGlnGlyAspGlySerHisLeuAlaValHisAspThrAlaIleGlyArgLeuGly 160  
QY 481 GCGCTGTGTGCTGGAGCAGCATCAGCATTTGTGAAATACGCGCATGTAGCGCGCGAG 540  
Db 161 AlaLeuYsCysTrpGlnHisIleGlnProLeuSerIlyTrAlaMetTrpAlaAlaAsp 180  
QY 541 GAACAGGTCCAGCTGCGGTGCGCGAGCTTACGCTTATTCGCGCGCGCTATGCG 600  
Db 181 GlnGlnValHisValAlaSerTrpProSerPheSerLeuTrpArgGlyMetAlaTrpAla 200  
QY 601 CTCGGAGCGAGGTTCATACCGCGCGAGCGAGCATTCACCGCGGTGCGAGCGGTCTAC 660  
Db 201 LeuGlyProGlnValAsnThrAlaAlaSerGlnIleTrpAlaValGlnGlyCysTrp 220  
QY 661 GTGCTGCGCTGCTGCGCGAGCGCTTCCCGGAGATGATCAAGATTTGATGATACCGC 720  
Db 221 ValLeuAlaSerCysAlaThrValSerProGlnMetIleIlySerValLeuValAspThrPro 240  
QY 721 GACAGGAGATGTTCTCAAGCGCGCGCGCGCTTTCATGATTTTCGCGCGCGAGCGC 780  
Db 241 AspIysGlnMetPheLeuYsAlaGlyGlyGlyPheAlaMetCilePheGlyProAspGly 260  
QY 781 CGGCGCGCGCGAGCGCGCTCCCGAGACGCGAGAGGAGCGGTGCGCGAGATGAC 840  
Db 261 ArgAlaLeuAlaGlnProLeuProGlnThrGlnGlnLeuValAlaAspIleAsp 280  
QY 841 CTCGGCATGATCGCTTGGCGCGAGCGCGCGCGATCCGCGCGCGCATTTTCAGCGCGC 900  
Db 281 LeuGlyMetIleAlaLeuAlaIlyAlaAlaAspProAlaGlyHisTrpSerArgPro 300  
QY 901 GACGTAAACGGCGCTGCTGCTGATGACGTCGCGCGCGCGCGCTGCTACGCTTATGATCC 960  
Db 301 AspValThrArgLeuLeuLeuAspArgArgProAlaGlnArgValValThrLeuAspAla 320  
QY 961 GCATTCGACACCGCAACGAGAGACAGAGCGAGCGCGCGCGCTGCTGCTGCTGCGGAA 1020  
Db 321 AlaPheGlnProGlnAsnGlnAspIlyAspAlaProAlaLeuArgValValAlaGln 340  
QY 1021 AGCGCGCGCGCGCGAG 1038  
Db 341 SerAlaAlaAlaAlaGln 346  
RESULT 7  
ID ADC23767 standard; protein; 341 AA.  
XX ADC23767;  
AC  
XX 18-DEC-2003 (first entry)  
DT  
XX Protein sequence (SeqID 34) exhibiting nitrilase activity.  
DE  
XX enzyme; nitrilase; nitrile; cyanohydrin; ammonia; biocatalyst;

KW enantiomer; chiral medicine.  
XX  
OS Unidentified.  
XX  
PN MO200300840-A2.  
XX  
PD 03-JUN-2003.  
XX  
PF 15-MAY-2002; 2002MO-US015983.  
XX  
PR 21-JUN-2001; 2001US-0300189P.  
XX  
PR 30-JUL-2001; 2001US-030906P.  
PR 22-JAN-2002; 2002US-0351336P.  
XX  
PA (DIVE-) DIVERSA CORP.  
PA (MADD/) MADDEN D.  
PI Madden M, Desautels G, Chaplin JA, Weiner DP, Milan A, Chi E;  
PI Short JM, Burk M;  
XX WPI; 2003-201417/19.  
XX DR N-PSDB; ADC23766.  
XX PT Novel nitrilase polypeptide, useful for making (R)- or (S)-ethyl-4-cyano-  
PT 3-hydroxybutyric acid or (R)- or (S)-mandelic acid or (S)- or (R)-phenyl  
PT lactic acid derivative and for producing pharmaceutical composition, and  
PT food additive.  
XX Claim 40; SEQ ID NO 34; 560pp; English.  
PS  
XX  
CC This invention relates to nitrilases and the nucleic acids that encode  
CC these enzymes thereof. Specifically, it refers to polypeptides that  
CC exhibit nitrilase activity, i.e. the ability to directly hydrolyse  
CC nitriles or cyanohydrins into their corresponding carboxylic acids and  
CC ammonia. Nitrilases have commercial utility as biocatalysts for use in  
CC the synthesis of enantiomerically pure aromatic and aliphatic amino  
CC acids, as well as hydroxy acids, which are important for the development  
CC of chiral medicines. Furthermore, the present invention describes  
CC nitrilases, isolated from mesophilic microorganisms, that have improved  
CC activity and stability at increased pH and temperature. They are also  
CC inexpensive, efficient catalysts, have broad substrate specificity and  
CC are capable of chiral differentiation. This polypeptide is a protein  
CC sequence that exhibits nitrilase activity of the invention.  
XX  
SQ Sequence 341 AA;  
Alignment Scores:  
Pred. No.: 1,08e-118 Length: 341  
Score: 1464.00 Matches: 272  
Percent Similarity: 90.1% Conservative: 28  
Best Local Similarity: 81.7% Mismatches: 33  
Query Match: 77.8% Indels: 0  
DB: 7 Gaps: 0  
US-09-751-299-1 (1-1041) x ADC23767 (1-341)  
QY 1 ATGTCCGAGCCCATGACGAGTATCGCGCGCGCTGCGAGCGCGCGGTGTTCTC 60  
Db 1 MetLeuSerProValThrGlnTrpArgAlaAlaValGlnAlaAlaProSerPheLeu 20  
QY 61 GATCTGACCGCGCAGTCCGAGAAAGCGATCGCTGATGAGAGCGCGCGCAAGCAGAC 120  
Db 21 AspLeuAspArgThrValGlnYsThrIleAlaIleIleGlnAlaAlaGlnGlnAsp 40  
QY 121 GTGCGCTGATCGATTCGAGAGACCTTGATTCGCGCGCTATTCCTTTGATATGAGCTG 180  
Db 41 ValArgLeuIleAlaPheProGlnThrTriPleProGlyTrpProLeuTriPleTrpLeu 60  
QY 181 GGGCGCGCGCTTGGGGCATGCGCTTGGTCCAGCGCTATTTCGAAATTCGCTCGTGGC 240  
Db 61 GlySerProAlaTrpGlyMetArgPheValGlnArgTrpPheGlnAseSerLeuValArg 80  
QY 241 GGCAGCAGCAGTGGAGCGCGCTGCGGAGTGGCGCGCGCGCGCGCATGCTGCTG 300

```
Db      |||...|||
81 GlySerIysGlnTrpAsnAlaIleAlaAspAlaAlaGlyGlnAlaGlyMetThrValVal
QY      301 GCCGGCTATAGCGAGCGCGCGCGCGCAAGCTTATATGAGCGAGCGCATTTCCGCCCC
Db      |||...|||
101 ValGlyPheSerGluArgAlaGlyGlySerLeuTrpMetGlyGlnAlaIlePheGlyPro
QY      361 GATGGCGATCTGATCGCGCGCGCGCGCAAGCTTAAAGCTTACCATCGGAGCGCAAGCTG
Db      |||...|||
121 GluGlyLeuLeuIleAlaAlaArgAlaGlyLeuLysProThrHisAlaGluArgThrVal
QY      421 TTGGCGAGGAGAGACGCGAGCGCATCTCGCGGTGACGATACCGGATCGCGCGCTCGGC
Db      |||...|||
141 PheGlyGluGlyAspGlySerHisIleuAlaValTyrGluThrGlyValGlyArgIleGly
QY      481 GCGCTCTGTTGCTGGAGAGCAATCCAGCATTTGCGAAATACGCCATGTACCGCGGAC
Db      |||...|||
161 AlaLeuCySerTrpGluHisIleGlnProLeuSerLysTyrAlaMetTyrAlaAlaAsn
QY      541 GAACAGGTCCAGTCTCGCGTCTGCGCGCGAGCTTCAAGCTTATCGCGGATGCTATGCG
Db      |||...|||
181 GluGlnValHisValAlaSerTrpProCysPheSerLeuTyrArgGlyMetAlaTyrAla
QY      601 CTGGAGCCGAGGATCATCCGCGCGAGCGCAAGCTTACGCGGTGCGAGCGCGCTGCTAC
Db      |||...|||
201 LeuGlyProGluValAsnThrAlaAlaSerGlnValTyrAlaValGluGlyCysTyr
QY      661 GTGCTGCGTGTGTCGCGGACCGTTCGCGCGAGATGATCAAGTATGTTGATGATACGCC
Db      |||...|||
221 ValLeuAlaSerCysLeuValValThrProGluIleLeuLysValLeuIleArgThrPro
QY      721 GACAGAGAGATGTTCTCAAGCGCGCGCGCTTTGCAATGATTTTCGGCGCGGAGCGC
Db      |||...|||
241 AspLysGluProLeuLeuAlaGlyGlyGlyPheSerMetIlePheGlyProAspGly
QY      781 CGCGCCCTGCGCGGACCGCTTCCGAGAGACCGAAGAGGAGTCTGTCGCGCATATCGAC
Db      |||...|||
261 ArgHisLeuAlaGlnProLeuProGluThrGluGluValThrAlaGluIleAsp
QY      841 CTCGGCGATGATCGGCTTGCGCAAGCGCGCGCGATCGCGCGCGCATTTTACCGCGCC
Db      |||...|||
281 LeuGlyAlaIleAlaLeuAlaLysAlaAlaAlaAspProAlaGlyHisTyrAlaArgPro
QY      901 GACGTAAACGCGGCTGCTGTGATCGACGTCGCCGCCCAACGCGTCTGACGCTTGATGCC
Db      |||...|||
301 AspValThrArgLeuLeuLeuAsnProArgProAlaAlaArgValAlaGluAlaLeuGlyPro
QY      961 GCATTGCAACGCAAAACGAGCAAGGAGGAGCGCGCC
Db      |||...|||
321 ArgPheGluValValGlnSerGluGlnAlaGluProPro
RESULT 8
ADH35868
ID      ADH35868 standard; protein; 341 AA.
XX      ADH35868;
AC      ADH35868;
DT      11-MAR-2004 (first entry)
XX      11-MAR-2004 (first entry)
DE      Chemical process monitoring; related nitrilase protein sequence SegID34.
XX      Chemical process monitoring; biochemical process monitoring; cyanide;
KW      high throughput system; enzyme.
XX      Unidentified.
OS      WO2003098187-A2.
XX      WO2003098187-A2.
PD      27-NOV-2003.
XX      27-NOV-2003.
PF      15-MAY-2003; 2003WO-US015639.
XX      15-MAY-2003; 2003WO-US015639.
PR      15-MAY-2002; 2002US-0380737P.
```

```
XX      (DIVE-) DIVERSA CORP.
PA      Weiner D, Chaplin JA, Chi E, Milan A, Desantis G, Burk MJ;
XX      Mcquaid J, Stege J;
PI      WPI, 2004-142708/14.
XX      N-PSDB; ADH35867.
DR      Monitoring a chemical or biochemical process comprises providing a
PT      reagent comprising a cyanide or a material that can be converted to
PT      cyanide or a reagent that generates a cyanide or a material that can be
XX      converted to cyanide.
PS      Claim 73; SEQ ID NO 34; 277pp; English.
XX      This invention relates to a novel method of monitoring chemical or
CC      biochemical processes. The method involves providing a reactant
CC      comprising cyanide (or a material that can be converted to a cyanide)
CC      that generates as a reaction product cyanide or a material that can be
CC      converted to cyanide and measuring the concentration of produced cyanide.
CC      The method is useful for monitoring a chemical or biochemical process.
CC      The method is effective for high throughput systems and is sufficiently
CC      sensitive to detect a small amount of product. The present sequence is
CC      that of a nitrilase protein which may be used in the method of the
XX      invention.
SQ      Sequence 341 AA;
XX      Alignment Scores:
Pred. No.: 1,08e-118 Length: 341
Score: 1464.00 Matches: 272
Percent Similarity: 90.1% Conservative: 28
Best Local Similarity: 91.7% Mismatches: 33
Query Match: 77.8% Indels: 0
DB: Gaps: 0
US-09-751-299-1 (1-1041) x ADH35868 (1-341)
QY      1 ATGTCGAGCCCATCGAAGATATCGGCGCGCGTGCAGAGCGCGCGGTGTTCTC
Db      |||...|||
1 MetLeuSerProValThrGlnTyrArgAlaAlaValGlnAlaAlaProSerPheLeu
QY      61 GATCTGACCGCGCAGTCTGAGAAAGGATCGGCTGATGAGACGCGCGCGCAAGCGAC
Db      |||...|||
21 AspLeuAspArgThrValGlnLysThrIleAlaIleIleGlnAlaAlaGluGlnAsp
QY      121 GTGCGCTGATGCGATCTCCAGAGACTTGATTCGCGGTATCCCTTTGGATATGCGTG
Db      |||...|||
41 ValArgLeuIleAlaPheProGluThrTrpIleProGlyTyrProLeuThrPleu
QY      181 GCGCGCGCGGCTTGCGGAGATGCGCTTCGTCGAGCGCTATTTGAGAAATTCGTCGCGCG
Db      |||...|||
61 GlySerProAlaTrpGlyMetArgPheValGlnArgTyrPheGluAsnSerLeuValArg
QY      241 GCGAGCAACAGTGGCGAGGCGCTGCGGATGCGGCGCGCGCGCGCACGATGATGCTG
Db      |||...|||
81 GlySerLysGlnTrpAsnAlaIleAlaAspAlaAlaArgHisAlaGlyMetThrValVal
QY      301 GCCGGCTATAGCGAGCGCGCGCGCGCAAGCTTATATGAGCGAGCGCATTTCCGCCCC
Db      |||...|||
101 ValGlyPheSerGluArgAlaGlyGlySerLeuTrpMetGlyGlnAlaIlePheGlyPro
QY      361 GATGGCGATCTGATCGCGCGCGCGCGCAAGCTTAAAGCTTACCATCGGAGCGCAAGCTG
Db      |||...|||
121 GluGlyLeuLeuIleAlaAlaArgAlaGlyLeuLysProThrHisAlaGluArgThrVal
QY      421 TTGGCGAGGAGAGACGCGAGCGCATCTCGCGGTGACGATACCGGATCGCGCGCTCGGC
Db      |||...|||
141 PheGlyGluGlyAspGlySerHisIleuAlaValTyrGluThrGlyValGlyArgIleGly
QY      481 GCGCTCTGTTGCTGGAGAGCAATCCAGCATTTGCGAAATACGCCATGTACCGCGGAC
```



Db 161 AlaleucCyseYstrpGluHisIleGlnProleuSerlystrYralametyYralaAlaAsn 180  
Qy 541 GAACAGGTCCACGTCGGGTGTCGGCCGAGCTTCAGCTCTATCCGGGATAGCTTACGG 600  
Db 181 GluGlnValHisValAlaSerTrpProCysePheSerleuYrargGlyMetAlaYrAla 200  
Qy 601 CTCGACCGGAGGTCAATACCGCCGAGACGAGATCTACGGGGTGAAGGGCGCTGTAC 660  
Db 201 LeuGlyProGluValAsnThrAlaAlaSerGlnValIlyrAlaGluGlyGlyCystrYr 220  
Qy 661 GTGTCGCGTCGTGCGGACCGCTCCCGAGATGATCAAGTATGATGATGATACGCC 720  
Db 221 ValLeuAlaSerCyseLeuValIValThrProGluIleuYsValIleuIleAspThrPro 240  
Qy 721 GACAAGAGATGTTCTTCAGAGCCGCGCGCGGTTTGCATGATTTTCGGGCCGACGCG 780  
Db 241 AspIyegIuProleuLeuAlaGlyGlyGlyPheSerMetIlePheGlyProAspGly 260  
Qy 781 CGGCGCTGCGCGGACCGCTCCCGAGAGACCGAGATGCTGCTGCGCATATGAC 840  
Db 261 ArgAlaLeuAlaGlnProleuProGluThrGluGluValIThrAlaGluIleAsp 280  
Qy 841 CTCGCGCATGATCGCGTGGCCAGAGCGCGCGCGCATCCGCGGCACTATTCAAGCGCC 900  
Db 281 LeuGlyAlaIleAlaLeuAlaIlyAlaAlaAspProAlaGlyHisYrAlaAspPro 300  
Qy 901 GACGTAAAGCGGCTGCTGCTGATGACGTCGCGGCCAGCGCTGTCACGCTTGATGCC 960  
Db 301 AspValIThrArgLeuLeuAlaAspProArgProAlaAlaArgValIAlaGluAlaGlyPro 320  
Qy 961 GCATTGCAACCGCAAAACGAGAGACAGAGGCGACCGCGCC 999  
Db 321 ArgPheGluValValGlnSerGluAlaGluPro 333

## RESULT 9

ID ADG93569 standard; protein; 341 AA.

AC ADG93569;

DT 11-MAR-2004 (first entry)

DE NitriIase enzyme amino acid sequence SeqID14.

KW nitriIase; nitrite; carboxylic acid; chemical process; pH; temperature;  
KM enantioselective transformation; enzyme.

XX Unidentified.

OS WO2003097810-A2.

PN 27-NOV-2003.

PD 15-MAY-2003; 2003WO-US015712.

PR 15-MAY-2002; 2002US-00146772.

PR 09-SEP-2002; 2002US-00241742.

PA (DIVE-) DIVERSA CORP.

PI Desantis G, Short JM, Burk M, Wong K, Farwell R, Chatman K;

DR WPI; 2004-090637/09.

DR N-PSDB; ADG93568.

XX New isolated or recombinant nucleic acid encoding a polypeptide having  
PT nitriIase activity, useful for screening enantioselective transformation.

XX Claim 44; SEQ ID NO 34; 295bp; English.

CC This invention is related to a novel isolated or recombinant nucleic acid  
encoding a protein having nitriIase activity. NitriIase's are capable of  
converting nitrite's directly to carboxylic acids and have great

CC potential for use in industrial chemical processes. The isolated  
CC nitriIase proteins of the invention have increased activity and stability  
CC at increased pH and temperature when compared to those conventionally  
CC used. In addition, the nucleic acid of the invention is useful for  
CC screening enantioselective transformation. The present sequence is that  
CC of a nitriIase enzyme of the invention.

XX Sequence 341 AA;

## Alignment Scores:

Pred. No.:	1,08e-118	Length:	341
Score:	1464.00	Matches:	272
Percent Similarity:	90.1%	Conservative:	28
Best Local Similarity:	81.7%	Mismatches:	33
Query Match:	8	Indels:	0
DB:	8	Gaps:	0

US-09-751-299-1 (1-1041) x ADG93569 (1-341)

Qy 1 ATGTGGAGCCCATGACGAAGTATCCGGCGCGCGGTGACGGCCGCGGTTCCTC 60  
Db 1 MetLeuSerProValIThrGlnYrArgAlaAlaValGlnAlaIleAlaProSerPheLeu 20  
Qy 61 GATCTGACCGCACGTCGAGAAAGCATCGCGCTGATCGAGAGCGCGCAAGAGAC 120  
Db 21 AspLeuAspArgThrValGlyIystrIleAlaIleIleGluGlnAlaGluGlnAsp 40  
Qy 121 GTGCGCTGATGCGATTCGCCAGAGCTGGATTCGCGCTATCCCTTTGATATGCGCTG 180  
Db 41 ValArgLeuIleAlaPheProGluThrTrpIleProGlyYrProleuTrpIleTrpLeu 60  
Qy 181 GGGCGCGCGCTGGGGCATGCGCTTCGTCACCGCTATTCGAGATTCGTCGCGCG 240  
Db 61 GlySerProAlaTrpGlyMetArgPheValGlnArgYrPheGluAsnSerLeuValArg 80  
Qy 241 GGCAGCAAGCATGCGCAGCGCTGCGGATGCGCGCGCGCGCATGATGTCGTG 300  
Db 81 GlySerIleGlnTrpAsnAlaIleAlaAspAlaAlaArgArgHisArgMetThrValVal 100  
Qy 301 GCCGCTATAGCAGCGCGCGCGCGCGCGCATCTATATGCGCGAGCGATCTTCGCGCC 360  
Db 101 ValGlyPheSerGluTrpAlaGlyGlySerLeuYrMetGlyGlnAlaIlePheGlyPro 120  
Qy 361 GATGCGCATGATTCGCGCGCGCGCGCGCGCATGACCTCAAGCTCCATCGCGAGCGCAC 420  
Db 121 GluGlyGluLeuIleAlaAlaArgGlyLeuIleProThrHisAlaGluArgThrVal 140  
Qy 421 TTGCGGAGGAGAGCAGCGCGCATCTCGCGGTGACAGTACCGCATCGGCGCGCTCGGC 480  
Db 141 PheGlyGluGlyAspGlySerHisIleAlaValAlaIytrGluThrGlyValGlyArgIleGly 160  
Qy 481 GCGCTCTGTGCTGGAGACATCCAGCATTCGCAATATGCAATATGATGACCGCGCAG 540  
Db 161 AlaleucCyseYstrpGluHisIleGlnProleuSerlystrYralametyYralaAlaAsn 180  
Qy 541 GAACAGGTCCACGTCGGGTGTCGGCCGAGCTTCAGCTCTATCCGGGATAGCTTACGG 600  
Db 181 GluGlnValHisValAlaSerTrpProCysePheSerleuYrargGlyMetAlaYrAla 200  
Qy 601 CTCGACCGGAGGTCAATACCGCCGAGACGAGATCTACGGGGTGAAGGGCGCTGTAC 660  
Db 201 LeuGlyProGluValAsnThrAlaAlaSerGlnValIlyrAlaGluGlyGlyCystrYr 220  
Qy 661 GTGTCGCGTCGTGCGGACCGCTCCCGAGATGATCAAGTATGATGATGATACGCC 720  
Db 221 ValLeuAlaSerCyseLeuValIValThrProGluIleuYsValIleuIleAspThrPro 240  
Qy 721 GACAAGAGATGTTCTTCAGAGCCGCGCGCGGTTTGCATGATTTTCGGGCCGACGCG 780  
Db 241 AspIyegIuProleuLeuAlaGlyGlyGlyPheSerMetIlePheGlyProAspGly 260  
Qy 781 CGGCGCTGCGCGGACCGCTCCCGAGAGACCGAGAGATGCTGCTGCGCATATGAC 840

DB 261 ArgAlaLeuAlaGlnProLeuProGluThrGluGluGlyLeuValThrAlaGluIleAsp 280

OY 841 CTCGGCATGATCGCGCTTGCGGCAAGCGCGCGCGCATCGCGCGGCCACTATTACGGCGCC 900

DB 281 LeuGlyAlaIleIleAlaLeuAlaIleAlaAlaIleAspProAlaGlyIleIleValArgPro 3000

OY 901 GACCGAAGCGGCGCTGCTGATGACGCTCGCGGCCACGCGCGTGCACGCTTGATGCC 960

DB 301 AspValThrArgLeuLeuLeuAsnProArgProAlaAlaArgValGluAlaLeuGlyPro 3200

OY 961 GCATTGCAACCGCAAAACGAGACAGCGCGCGCGCGCC 999

DB 321 ArgPheGluValValGlnSerGluGlnAlaGluProPro 333

RESULT 10

ADI62166

ID ADI62166 standard; protein; 341 AA.

XX ADI62166;

XX AC

XX 22-APR-2004 (first entry)

DT Nicotianase polypeptide #17.

DE

XX Atorvastatin; (R)-ethyl 4-cyano-3-hydroxybutyrate;

XX (R)-ethyl 4-cyano-3-hydroxybutyric acid; epichlorohydrin;

XX 3-hydroxyglutaronitrile; 3-hydroxyglutaronitrile;

XX 4-cyano-3-hydroxybutyric acid; ethyl-4-cyano-3-hydroxybutyric acid;

XX mixed hyperlipidaemia; homozygous familial hypercholesterolaemia;

XX antilipaeitic; enzyme.

XX Unidentified.

XX OS

XX MO2003106415-A2.

PN

PD 24-DEC-2003.

PF 13-JUN-2003; 2003MO-US018840.

PR 13-JUN-2002; 2002US-0389317P.

PR 28-JUN-2002; 2002US-0392944P.

XX

PA (DIVE-) DIVERSA CORP.

XX

P1 Burk M, Desantis G, Morgan B, Zhu Z;

XX

XX WPI; 2004-090821/09.

DR N-PSDB; ADI62165.

XX

XX

PT Preparation of atorvastatin comprises catalytic conversion of 3-

PT hydroxyglutaronitrile by polypeptide with nitrilase activity, converting

PT obtained 4-cyano-3-hydroxybutyric acid to ethyl-4-cyano-3-hydroxybutyric

PT acid and forming atorvastatin.

XX

PS Claim 41; SEQ ID NO 34; 253pp; English.

XX

XX The present invention relates to a method for preparing an atorvastatin

XX intermediate known as (R)-ethyl 4-cyano-3-hydroxybutyrate ((R)-ethyl 4-

XX cyano-3-hydroxybutyric acid). The method comprises optionally converting

XX epichlorohydrin or equivalent to 3-hydroxyglutaronitrile, catalytic

XX conversion of 3-hydroxyglutaronitrile or equivalent to 4-cyano-3-

XX hydroxybutyric acid with a polypeptide having nitrilase activity,

XX converting 4-cyano-3-hydroxybutyric to ethyl-4-cyano-3-hydroxybutyric

XX acid, and converting this to (R)-ethyl 4-cyano-3-hydroxybutyric

XX method involves whole cell processes, cell lysate process, "one pot"

XX processes, and "multi-pot" processes using a variety of parameters.

XX Atorvastatin is used, in conjunction with dietary restriction, in the

XX management of hyperlipidaemia, including hypercholesterolaemia, mixed

XX dyslipidaemia and homozygous familial hypercholesterolaemia. The present

XX sequence represents a nitrilase polypeptide obtained from an

XX environmental sample.

XX

XX Sequence 341 AA:

XX

Alignment Scores:		1,08e-118	Length:	341
Pred. No.:	1464.00	Matches:	272	
Score:	90.1%	Conservative:	28	
Percent Similarity:	81.7%	Mismatches:	33	
Best Local Similarity:	77.8%	Indels:	0	
Query Match:	8	Gaps:	0	
US-09-751-299-1 (1-1041) x AD162166 (1-341)				
QY	1	ATGTGGAGCCCATGACGAGATATCGGCGCGCGGTGACAGCCCGCGATGTTCTC	60	
DB	1	MetluseerProvalThrgintYrAglalalAalValGlnalalAProSerPheleu	20	
QY	61	GATTCGACCGGCACAGTCGAGAGAGAGATCGGCGCTGATGACAGCGGCCACAGGAC	120	
DB	21	AsplenuspbrgThrValGlnYstrHrleAlailelleglGlnalalAglucInasp	40	
QY	121	GTGGCGCTGATTCGATTTCCAGAGACTTGATTCGCGGATACCTTTTGATATGCTG	180	
DB	41	ValargleuilealAphneprogiutHrTripleprogiYrProleutHrTripleu	60	
QY	181	GGCGCGCGCGCTTGGGGCATGCGCTTCCTCCAGCGCTATTTCGAGATTCCGTGCGC	240	
DB	61	GlyserProalTrpGlymetAargPheValGlnAArgYrPheGlnAsnSerleuValArg	80	
QY	241	GGCGAGCAAGAGTGGAGCGCGCTGGGCGAGTGGCGCGCCGCCACGCGATCATGTGCG	300	
DB	81	GlyserYleGlnTrpAsnAlailealAaPheAlaAlaArgAghlAArgmetHrValVal	100	
QY	301	GCGCGCATATGCGAGCGCGCGCGCGCGAGCTCTATATAGGCGACGCGATTTGGCCCC	360	
DB	101	ValGlyPheSerleuAargAlaGlyGlySerleuYrMetGlyGlnalallePheGlyPro	120	
QY	361	GATGCGCATGTGATTCGCGCGCGCGCGCGCAAGCTCAAGCTTACCATTCGCGAGCGACGCGT	420	
DB	121	GlnGlyleuileuilealAalAaArgArgYleuYrProThrHsAlaGlnAargThrVal	140	
QY	421	TTTGGCGAGGGAAGACGCGACCATCTCGCGGTGACGATCCGCGCATCGGCGCTTCGCG	480	
DB	141	PheGlyleuGlyAspGlySerHsIleuAlaValYrGlnTrpGlyValGlyArgIleGly	160	
QY	481	GCGCTGTGTTGCTGGGAGACATTCAGCGCATTTGTGAAATACGCGATGACCGCGCGAC	540	
DB	161	AlaIeucYseYrTrpGlnHsIleGlnProIeucSerYrIYrAlaMetYrAlaAlaAsn	180	
QY	541	GAACAGGTCCAGCTCGCGTGTGGCGCGAGCTTCATTCGCGGCGATGCGCTATGCG	600	
DB	181	GlnGlnValHsIleValAlaSerTrpProCysPheSerleuYrAArgIlymetAlaYrAla	200	
QY	601	CTGCGACCGGAGGTCAATACCGCGCGCAAGCATCTACGGGTCGAGGGCGCGCTGAC	660	
DB	201	LeuGlyProGlnValAsnThrAlaSerGlnValYrAlaValGlnGlyGlyCysYr	220	
QY	661	GTGCTGCGCTGTGGCGCGACCGTTTCGCGCGAGATGATCAAGATATTGTGATACGCC	720	
DB	221	ValleuAlaSerCysleuValAlaThrProGlnIleleuYrValleuIleAepThrPro	240	
QY	721	GACAAAGAGATGTTCTCAAGCGCGCGCGCGGTTTTCGAGATTTTCGCGCGCGACGC	780	
DB	241	AspYleGlnProleuIleuAlaGlyYrPheSerMetIlePheGlyProAaGly	260	
QY	781	CGCGCGCTGAGCGCGCGCTCCCGGAGACGGAAGAGGACTGTCGTGCGCGCATTCGAC	840	
DB	261	ArgAlaIleuAlaGlnProleuProGlnTrpGlnGlyleuValThrAlaGlnIleAaP	280	
QY	841	CTGGGATATGCGCTTGGCGCAAGCGCGCGCGCATTCGCGGGCGCATTTACGCGCCC	900	
DB	281	LeuGlyAlaIleAlaIleuAlaYsaAlaAlaAspProAlaGlyHsIYrAlaAlaArgPro	300	
QY	901	GAGTAAACCGCGCTGCTGTGATGAGATCGCGGCCAACGCGCTGCACGTTGATGCGC	960	

Db 301 AspvAlrThrArgLeuLeuLeuAsnProArgProAlaAlaArgValGluAlaLeuGlyPro 320  
QY 961 GCATTGACCGCGAAACGAGGACAGAGCGCGCCGCC 999  
Db 321 ArgPheGluValValGlnSerGluGlnAlaGluProPro 333  
RESULT 11  
ADI64287  
ID ADI64287 standard; protein; 341 AA.  
XX  
AC ADI64287;  
XX  
DT 22-APR-2004 (first entry)  
XX  
DE Nitriase seq id 18.  
XX  
XX (R)-ethyl 4-cyano-3-hydroxybutyric acid; nitrile hydrolysis;  
KM carboxylic acid; cyanohydrin moiety hydrolysis;  
KM ammonitrile moiety hydrolysis; chiral alpha-hydroxy acid molecule;  
KM chiral amino acid molecule; (S)-ethyl 4-cyano-3-hydroxybutyric acid;  
KM (R)-mandelic acid; (S)-mandelic acid; (S)-phenyl lactic acid derivative;  
KM (R)-phenyl lactic acid derivative; & enantiomeric excess;  
KM & diastomeric excess; food additive; drug intermediate; nitriase.  
XX  
OS Unidentified.  
XX  
XX US2004014195-A1.  
XX  
XX 22-JAN-2004.  
XX  
XX 15-MAY-2003; 2003US-00440523.  
XX  
XX 22-DEC-1999; 99US-0173609P.  
XX 07-DEC-2000; 2000US-0254414P.  
XX 28-DEC-2000; 2000US-00751299.  
XX 21-JUN-2001; 2001US-0300189P.  
XX 30-JUL-2001; 2001US-0309006P.  
XX 22-JAN-2002; 2002US-0351336P.  
XX 15-MAY-2002; 2002US-00146772.  
XX 09-SEP-2002; 2002US-00241742.  
XX  
XX (DIVE-) DIVERSA CORP.  
XX  
XX Desantis G, Short JM, Burk MJ, Wong K, Farwell R, Chatman K;  
XX WPI; 2004-121569/12.  
XX DR N-PSDB; ADI64286.  
XX  
XX Novel isolated or recombinant polypeptide having nitriase activity,  
XX PT useful in production of food additives.  
XX  
XX  
XX Claim 46; SEQ ID NO 34; 105pp; English.  
XX  
XX The invention describes an isolated or recombinant polypeptide (I)  
XX comprising amino acids having a sequence at least 50 % identical to a  
XX sequence (SI) available in electronic form (EC) from the following web  
XX site ftp.seagate.uspto.gov/sequence.html?DocID=2004014195, or its  
XX variants, having one or more mutations at residue 55 Lys, Gly or Glu, at  
XX residue 60 glutamic acid, at residue 111 Ser, their combinations or  
XX fragments. (I) is useful for: producing an (R)-ethyl 4-cyano-3-  
XX hydroxybutyric acid; hydrolysing a nitrile to a carboxylic acid;  
XX hydrolysing cyanohydrin moiety or an ammonitrile moiety; producing a  
XX chiral alpha-hydroxy acid molecule or a chiral amino acid molecule;  
XX producing an (S)-ethyl 4-cyano-3-hydroxybutyric acid; producing (R)-  
XX mandelic acid or (S)-mandelic acid; producing (S)-phenyl lactic acid  
XX derivative or an (R)-phenyl lactic acid derivative; modifying a molecule  
XX for and for identifying a modified compound. The inventive method is useful  
XX for monitoring or determining & enantiomeric excess or & diastomeric  
XX excess. (I) is useful in the production of food additives and drug  
XX intermediates. This is the amino acid sequence of a nitriase of the  
XX invention.  
XX  
XX Sequence 341 AA;

Alignment Scores:  
Pred. No.: 1.08e-118 Length: 341  
Score: 1464.00 Matches: 272  
Percent Similarity: 90.1% Conservative: 28  
Best Local Similarity: 81.7% Mismatches: 33  
Query Match: 77.8% Indels: 0  
DB: 8 Gaps: 0  
US-09-751-299-1 (1-1041) x ADI64287 (1-341)  
QY 1 ATGTCCGAGCCCATGACGAATATCCGCGCGCGCGTGCAGAGCCGCGGTGTTCTC 60  
Db 1 MetLeuSerProValThrGlnTyrArgAlaAlaValGlnAlaIleAlaProSerPheLeu 20  
QY 61 GATCTGACCGCGACAGTCCGAGAAAGCGATCGCGCTGATGCAGAGCGCGCAAGAGAC 120  
Db 21 AspvLeuAspArgThrValGluSerThrIleAlaIleIleGluGlnAlaIleGluGlnAsp 40  
QY 121 GTGGCGCTGATGCGATTCGCCAGACTGTGATTCGCCGCTATCCCTTTGGATATGCGCTG 180  
Db 41 ValArgLeuIleAlaPheProGluThrTyrIleProGlyTyrProLeuThrIleTyrPleu 60  
QY 181 GCGCGCCGCGCTTGGGCGCATGCGCTTCGTCACAGCGCTATTCGAGAATTGCGTGC 240  
Db 61 GlySerProAlaTrpGlyMetArgPheValGlnArgTyrPheGluLysSerLeuValArg 80  
QY 241 GGCAGAGACGATGGAGAGCGCGCTGCGGATGCGCGCGCGCGCGCATGATGTCGTG 300  
Db 81 GlySerIleGlnTrpPheAlaIleAlaPheAlaAlaArgArgIleSerGlyThrValVal 100  
QY 301 GCGCGCTATATGCGAGCGCGCGCGCGCGCGCTTATATGCGCGCGCGCGCGCGCGCG 360  
Db 101 ValGlyPheSerIleLysArgAlaGlyGlySerLeuTyrMetGlyGlnAlaIlePheGlyPro 120  
QY 361 GATGCGCATGATTCGCCG 420  
Db 121 GlnGlyGluLeuIleAlaAlaArgAlaGlyLeuLysProThrIleAlaGluGlnThrVal 140  
QY 421 TTGCGGAGAGAGACG 480  
Db 141 PheGlyGluGlyIlePheGlySerIleAlaValIleValIleValIleGlyAlaGlyArgIleGly 160  
QY 481 GCGCTCTGTTGCTGCGAGACATCCAGCCATTCGAAATAGCCCATGATCGCGCGCG 540  
Db 161 AlaLeuCySerTrpGluIleAlaIleGlnProLeuSerLeuTyrAlaMetTyrAlaAlaAsn 180  
QY 541 GAACAGGTCACG 600  
Db 181 GluGlnValHisValAlaSerTrpProCyPheSerLeuTyrArgGlyMetAlaTyrAla 200  
QY 601 CTGCGACCGGAGGTCAATACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTG 660  
Db 201 LeuGlyProGluValAlaThrAlaAlaSerGlnValTyrAlaValGluGlyCysTyr 220  
QY 661 GTCTGCGCGTCTGCG 720  
Db 221 ValLeuAlaSerCysLeuValValThrProGluIleLeuLysValLeuIleAspThrPro 240  
QY 721 GACAGAGAGATGTTCTCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 780  
Db 241 AspvLysGluProLeuLeuLeuAlaGlyGlyGlyPheSerMetIlePheGlyProAspGly 260  
QY 781 CCG 840  
Db 261 ArgAlaLeuAlaGlnProLeuProGluThrGluGluGlyLeuValThrAlaGluIleAsp 280  
QY 841 CTGCGCATGATGCGGTGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 900  
Db 281 LeuGlyAlaIleAlaLeuAlaLysAlaAlaAspProAlaIleIleTyrAlaAspPro 300  
QY 901 GACGTACGCGGCGTGTGCTGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 960  
Sequence 341 AA;

DB 301 AspValThrArgLeuLeuLeuAsnProArgProAlaAlaArgValGluValLeuGlyPro 320  
 QY 961 GCATTTCGAAACCCGAAACGAGACAGGCGCGGCC 999  
 DB 321 ArgPheGluValValGlnSerGluGlnAlaGluProPro 333  
 RESULT 12  
 ADC23749  
 ID ADC23749 standard; protein, 348 AA.  
 XX  
 AC ADC23749;  
 XX  
 DT 18-DEC-2003 (first entry)  
 XX  
 DE Protein sequence (SeqID 16) exhibiting nitrilase activity.  
 XX  
 KM enzyme; nitrilase; nitrile; cyanohydrin; ammonia; biocatalyst;  
 KM enantiomer; chiral medicine.  
 XX  
 OS Unidentified.  
 XX  
 PN W0203000840-A2.  
 XX  
 PD 03-JAN-2003.  
 XX  
 PF 15-MAY-2002; 2002WO-US015983.  
 XX  
 PR 21-JUN-2001; 2001US-0300189P.  
 PR 30-JUL-2001; 2001US-0309006P.  
 PR 22-JAN-2002; 2002US-0351336P.  
 XX  
 PA (DIVE-) DIVERSA CORP.  
 PA (MADN/) MADSEN D.  
 XX  
 PI Madden M, Desantis G, Chaplin JA, Weiner DP, Milan A, Chi E,  
 PI Short JM, Burk M;  
 DR N-PSDB; ADC23748.  
 DR WPI; 2003-201417/19.  
 PT Novel nitrilase polypeptide, useful for making (R)- or (S)-ethyl-4-cyano-  
 PT 3-hydroxybutyric acid or (R)- or (S)-mandelic acid or (S)- or (R)-phenyl  
 PT lactic acid derivative and for producing pharmaceutical composition, and  
 PT food additive.  
 XX  
 PS Claim 40; SEQ ID NO 16; 560pp; English.  
 XX  
 CC This invention relates to nitrilases and the nucleic acids that encode  
 CC these enzymes thereof. Specifically, it refers to polypeptides that  
 CC exhibit nitrilase activity, i.e. the ability to directly hydrolyse  
 CC nitriles or cyanohydrins into their corresponding carboxylic acids and  
 CC ammonia. Nitrilases have commercial utility as biocatalysts for use in  
 CC the synthesis of enantiomerically pure aromatic and aliphatic amino  
 CC acids, as well as hydroxy acids, which are important for the development  
 CC of chiral medicines. Furthermore, the present invention describes  
 CC nitrilases, isolated from mesophilic microorganisms, that have improved  
 CC activity and stability at increased pH and temperature. They are also  
 CC inexpensive, efficient catalysts, have broad substrate specificity and  
 CC are capable of chiral differentiation. This polypeptide is a protein  
 CC sequence that exhibits nitrilase activity of the invention.  
 XX  
 SQ Sequence 348 AA;  
 Alignment Scores:  
 Pred. No.: 1,97e-96 Length: 348  
 Score: 1209.00 Matches: 230  
 Percent Similarity: 79.3% Conservative: 42  
 Best Local Similarity: 67.1% Mismatches: 69  
 Query Match: 64.3% Indels: 2  
 DB: 7 Gaps: 2  
 US-09-751-299-1 (1-1041) x ADC23749 (1-348)

QY 10 CCATGACGAAG--TATCGCGCGCGCGGTGACAGCGCGCGGTTCTCGATCTC 66  
 DB 2 ProThSerIySgInPheArgValAlaAlaValGlnAlaAlaProValPheLeuAspLeu 21  
 QY 67 GACCGCACATCGAAGAAAGCGATCGGCTGATCGAGCGGCGGCAAGCAGACGTCGCG 126  
 DB 22 GluGlyAlaIleSerIySgIyIleSerLeuIleGluGluAlaIleSerAengIyAlaIy 41  
 QY 127 CTGATCGCATTCGACGAGACTTGATTCCTGATTCCTTGGATTCGATTCGCGCGCG 186  
 DB 42 LeuIleAlaPheProGluIleProGluIleProGluIleProGluIleProGluIlePhe 61  
 QY 187 CCGGCTTGGGCGATGCGCTTCGTCAGCGCTATTCAGAAATTCGCTGTCGCGCGAGC 246  
 DB 62 ProAlaIleProIyMetArgPheValGlnArgIyPheAspAsnSerIleuMetLeuGlySer 81  
 QY 247 AACGATGGCAGGCGCTGCGGATGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 306  
 DB 82 GluGlnAlaIySargMetAsnGlnAlaAlaAlaAsnAsnIyIleTyValValMetGly 101  
 QY 307 TATAGCAGCGCGCGCGCGCGAGCTCTATATGCGCGAGCGCATTCGCGCGCGATGCG 366  
 DB 102 TySerGluArgSerGlyIySerLeuIyMetGlyGlnSerIleIleAsnAspIySgIy 121  
 QY 367 GATCGATCGCGCGCGCGCGCGCAAGCTCAAGCTTACCATGCGGAGCGCACCGTTCGCG 426  
 DB 122 GluThrIlePheThrArgArgIySleuIySProThrIleValGluArgThrValPheGly 141  
 QY 427 GAGGAGACGCGCGCGCGCATTCGCGGTCAGCATACCGCGCGCGCGCGCGCGCGCG 486  
 DB 142 GluGlyAspGlySerIleuIySValMetAspThrGluIleGlyArgValGlyAlaMet 161  
 QY 487 TGTTCCTGGAGACATCCAGCCATTCGAAATACCGCATGTACCGCGCGAGCAAG 546  
 DB 162 CySCTyPgluIleuIleuGlnProIleuSerIyTyAlaMetTySerGlnAspGluGln 181  
 QY 547 GTCCAGCTCGCGCGCGCGCGAGCTTCAGCTTCATCGCGGCGCATGCGCGCGCGCG 606  
 DB 182 IleIleIleAlaSerIleProSerPheSerIleIyArgIyAlaAlaIyAlaIleuGly 201  
 QY 607 CCGGAGTCATATCCGCGCGCAAGCGCATCTACGCGGTGAGAGCGCGCGCTACGTCGTG 666  
 DB 202 ProGluLeuAsnAsnAlaIleSerGlnMetTyAlaAlaGluGlyGlnCySProValLeu 221  
 QY 667 GCGTGTGGCGCGCGCTTCGCGCGAGATGATCAAGATTCGTGATACGCCCGCAAG 726  
 DB 222 AlaProCyAlaIleThrValSerIySgIleuMetIleGluMetLeuIleAspAspProArgIyS 241  
 QY 727 GAGATGTTCTCAAGCGCGCGCGGTTTGGCATGATTTTGGCGCGCGCGCGCGCGCC 786  
 DB 242 GluProLeuLeuIleuGluGlyIyGlyIyPheThrMetIleTyGlyProAspIyArgPro 261  
 QY 787 CTGGCGGACCGCTCCCGGAGACCGAAGAGGATCTGTCGCGCATTCGACTTCGCGC 846  
 DB 262 LeuAlaIySProIleuProIleuAsnGluGluIyLeuIleTyAlaAspIleAspLeuGly 281  
 QY 847 ATGATGCGCTTGGCGAAGCGCGCGCGCGCGATTCGCGGCGCGCATTCGAGCGCGCA 906  
 DB 282 MetIleSerMetAlaIyAlaAlaAlaAspProAlaGlyIleTyAlaArgProAspVal 301  
 QY 907 ACGCGCTGCTCTGATGACGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGATT 966  
 DB 302 ThrArgLeuLeuPheAsnSerIleProAlaAsnArgValGluTyIleAsnProIleAsp 321  
 QY 967 GAACCGCAAAACGAGACAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1026  
 DB 322 GlyPro---ThGluSerLeuIyAspMetGlyIySValMetGluAlaIleGluGlnGln 340  
 QY 1027 GCCGCGCGCG 1035  
 DB 341 IySAlaAla 343  
 RESULT 13



XX MO2003097810-A2.  
XX 27-NOV-2003.  
XX 15-MAY-2003; 2003WO-US015712.  
XX 15-MAY-2002; 2002US-00146772.  
XX 09-SEP-2002; 2002US-00241742.  
XX (DIVE-) DIVERSA CORP.  
XX Desantis G, Short JM, Burk M, Wong K, Farwell R, Chatman K;  
XX WPI: 2004-090637/09.  
XX N-PSDB; ADG93550.  
XX New isolated or recombinant nucleic acid encoding a polypeptide having  
XX nitrilase activity, useful for screening enantioselective transformation.  
XX Claim 44; SEQ ID NO 16; 295bp; English.  
XX This invention is related to a novel isolated or recombinant nucleic acid  
XX encoding a protein having nitrilase activity. Nitrilase's are capable of  
XX converting nitrile's directly to carboxylic acids and have great  
XX potential for use in industrial chemical processes. The isolated  
XX nitrilase proteins of the invention have increased activity and stability  
XX at increased pH and temperature when compared to those conventionally  
XX used. In addition, the nucleic acid of the invention is useful for  
XX screening enantioselective transformation. The present sequence is that  
XX of a nitrilase enzyme of the invention.  
XX Sequence 348 AA;  
XX

Alignment Scores:  
Pred. No.: 1 97e-96 Length: 348  
Score: 1209.00 Matches: 230  
Percent Similarity: 79.3% Conservative: 42  
Best Local Similarity: 67.1% Mismatches: 69  
Query Match: 64.3% Indels: 2  
Gaps: 2

US-09-751-299-1 (1-1041) x ADG93551 (1-348)

QY 10 CCCATGACGAG--TATGCGGCGGCGGCGGTCAGAGCCGCGGTTCCTCATTC 66  
DB 2 ProthSerIyglInphearYValAlaValAlaGlnAlaAlaProValPheLeuAspLeu 21  
QY 67 GACCGCAGAGTCGAGAAAGCATGCGGCTGATCGAGCAGGCGGCAAGCAGCAGTGGCG 126  
DB 22 GIUGIYAlaIleSerIySglIleSerLeuIleGIUGIAlaAlaSerInglYAlaYs 41  
QY 127 CTGATGCAATCCAGAGACTTGATTCGCGCTATCCCTTTGGATATGCTGGCGCG 186  
DB 42 LeuIleAlaPheProGIuThrIleProGIYTYProItrIleTrpIleLeuAspSer 61  
QY 187 CCGGCTTGGGAGATGCGCTTCGTCAGAGCTATTTGAGAAATTCGTCGCGGCGGAGC 246  
DB 62 ProAlaItrGIYmeArIySerIyValGIuIyGTYrPheAspAsnSerIleuMetLeuGIYSer 81  
QY 247 AAGCAGTGGCAGGCGCTGGCGAGTGGCGCGCGCGGCAAGCAGATGTCGTGGCGGC 306  
DB 82 GIUGIAlaIaIyArIyglmetAengIAlaAlaAlaAsnAsnIyIleTYrValValmetGIY 101  
QY 307 TATAGCAGCGCGCGCGGCGGAGCTTATGAGGCGAGCTTCCTCGCGCGCGATGGC 366  
DB 102 IYSerGIuYrSglIySerIyIleYrmeGIYelInSerIleIleAsnAspIySglY 121  
QY 367 GATCTGATCGCGCGCGGCGGCAAGCTCAAGCTTACCATCGGAGCGCAACCGTTCGCG 426  
DB 122 GIuThrIlePheThrIyArIySglYIleuIySglYProThIleValGIuIyThrValPheGIY 141  
QY 427 GAGGAGACGCGGAGCATCTTCGCGGTGCAGATACCGCCATCGGCGCGCTC 486

DB 142 GIUGIYAspGIYSerIleIleuCYsValIleAspIhrGIuIleGIYArValGIYAlaMet 161  
QY 487 TGTTCCTGGAGAGCATTCAGACCTTGTCCGAAATTCAGCATGTACCGCGCGGAGCAAG 546  
DB 162 CYsCYsTrpGIuIleSglInProLeuSerIyTYrAlaMetIYrSerGIuAspGIuGIu 181  
QY 547 GTCCAGCTGCGGTGCTGGCGAGCTTACGCTTATCGCGGAGCATGAGCTTATCGCTCGGA 606  
DB 182 IleHISleIleAsnSerIleProSerPheSerIleuYrIyrgIYAlaIaIyAlaIleuGIY 201  
QY 607 CCGAGGCTCAATACCGCGCGGCAAGCATCTACCGCGGTTCAGAGGCGCGCTGTAGTGTG 666  
DB 202 ProGIuIleuAsnAlaIleAsnSerGIuMetIYrAlaIaGIuGIYGIuCYsPheValIleu 221  
QY 667 GCGTGTGCGGCGGCGGCTTCGCGGAGATGATCAAGATTTGTGATATACCGCGGACAG 726  
DB 222 AlaProCYsAlaItrIyValSerIySglIuMetIleGIuMetLeuIleAspAspProAlaYs 241  
QY 727 GAGATGTTCTTCMAAGCGCGGCGGCTTTCGATGATTTTCGCGCGCGAGCGCGCGC 786  
DB 242 GIuProLeuIleuGIuGIYGIYIYpHeThrMetIleTYrGIYProAspGIYArPro 261  
QY 787 CTGGCGGAGCGGCTTCGCGGAGAGAGGAGCATGTGTCGCGCGATATACCTCGGC 846  
DB 262 LeuAlaIySProLeuProGIuAsnGIuGIuIleuIleuTYrAlaIaAspIleAspLeuGIY 281  
QY 847 ATGATGCGGTTGGCCAGGCGGCGCGCATTCGCGGCGGCACTATTCAGCGCGCGAGTA 906  
DB 282 MetIleSerMetAlaIyAlaAlaIaAspProAlaGIYIleIYrAlaIyArProAspVal 301  
QY 907 ACGCGGCTGCTGCTGATTCAGATCGTCGCGCGCAAGCGGTTCAGCGCGCATTC 966  
DB 302 ThrArgLeuIleuPheAsnSerAlaProAlaAsnArgValGIuTYrIleAsnProAlaSer 321  
QY 967 GAACCGCAAAACGAGAGCAGGCGCGCGCGCGCTGCGGTGTGTGGAGAAAGCGCC 1026  
DB 322 GIYPro---ThrGIuSerLeuIyAspMetGIYsMetGIuMetGIuAlaGIuGIuGIu 340  
QY 1027 GCCGCGCGG 1035  
DB 341 IySAlaAla 343  
RESULT 15  
ADI62148  
ID ADI62148 standard; protein; 348 AA.  
XX  
AC ADI62148;  
XX  
DT 22-APR-2004 (first entry)  
XX  
DE Nitrilase polypeptide #8.  
XX  
XX Alorvaetatin; (R)-ethyl 4-cyano-3-hydroxybutyrate;  
XX (R)-ethyl 4-cyano-3-hydroxybutyric acid; epichlorohydrin;  
XX 3-hydroxyglutaronitrile; 3-hydroxyglutaronitrile;  
XX 4-cyano-3-hydroxybutyric acid; ethyl-4-cyano-3-hydroxybutyric acid;  
XX mixed hyperlipidaemia; homozygous familial hypercholesterolaemia;  
XX antilipemic; enzyme.  
XX  
OS Unidentified.  
XX  
PN WO2003106415-A2.  
XX  
XX  
PD 24-DEC-2003.  
XX  
PF 13-JUN-2003; 2003WO-US018840.  
XX  
XX 13-JUN-2002; 2002US-0389117P.  
PR 28-JUN-2002; 2002US-0392944P.  
XX  
XX  
PA (DIVE-) DIVERSA CORP.

PI Burk M, Desantis G, Morgan B, Zhu Z;  
XX WPI: 2004-090821/09.  
DR N-PSDB; ADI62147.

XX Preparation of atorvastatin comprises catalytic conversion of 3-hydroxyglutaronitrile by polypeptide with nitrilase activity, converting obtained 4-cyano-3-hydroxybutyric acid to ethyl-4-cyano-3-hydroxybutyric acid and forming atorvastatin.

XX Claim 41; SEQ ID NO 16; 253bp; English.

XX The present invention relates to a method for preparing an atorvastatin intermediate known as (R)-ethyl 4-cyano-3-hydroxybutyrate ((R)-ethyl 4-cyano-3-hydroxybutyric acid). The method comprises optionally converting epichlorohydrin or equivalent to 3-hydroxyglutaronitrile, catalytic conversion of 3-hydroxyglutaronitrile or equivalent to 4-cyano-3-hydroxybutyric acid with a polypeptide having nitrilase activity, converting 4-cyano-3-hydroxybutyric acid to ethyl-4-cyano-3-hydroxybutyric acid, and converting this to (R)-ethyl 4-cyano-3-hydroxybutyrate. The method involves whole cell processes, cell lysate process, "one pot" processes, and "multi-pot" processes using a variety of parameters. Atorvastatin is used, in conjunction with dietary restriction, in the management of hyperlipidaemia, including hypercholesterolaemia, mixed dyslipidaemia and homozygous familial hypercholesterolaemia. The present sequence represents a nitrilase polypeptide obtained from an environmental sample.

XX Sequence 348 AA;

Alignment Scores:

Pred. No.:	1,97e-96	Length:	348
Score:	1209.00	Matches:	230
Percent Similarity:	79.3%	Conservative:	42
Best Local Similarity:	67.1%	Mismatches:	69
Query Match:	64.3%	Indels:	2
DB:	8	Gaps:	2

US-09-751-299-1 (1-1041) x ADI62148 (1-348)

QY 10 CCCATGACGAAG--TATCGCGGCGCGGATGCGAGCCGCGGCTGTTCTCGATCTC 66  
DB 2 ProThSerYsgInPheArgValAlaAlaValGlnAlaIleProValPheLeuAspLeu 21  
QY 67 GACCGCACAGTCCGAAAGGAGTGGCTGATGACGAGCGGCGCAAGCAAGAGTGGCC 126  
DB 22 GlnGlyAlaIleSerYsgIleSerLeuIleGlnGlnAlaIleSerAenGlyAlaIle 41  
QY 127 CTGATGCGATTCCTCCAGAGACTTGATTCCTCCGCTATCCCTTTGATATGCTGGGCGG 186  
DB 42 LeuIleAlaPheProGlnTrpIleProGlyTrpProTrpIleTrpIleAspSer 61  
QY 187 CCGGCTTGGGAGATCGCTTCGTCAGCGCTATTCGAGAAATTCGTCGCGCGGAGC 246  
DB 62 ProIleAlaTrpGlnMetArgPheValGlnArgTrpPheAspAsnSerLeuMetLeuGlySer 81  
QY 247 AAGCAGTGGCAGCGCCCTGCGGAGTGGCGCCCGCGCCAGCATGATGTCGTCGCGG 306  
DB 82 GlnGlnAlaIleAspArgMetAenGlnAlaIleAlaAsnAsnIleTyrValValMetGly 101  
QY 307 TATAGCGAGCGCGCGCGCGCGCGCTTATATAGCGCGAGCGAGCTTCGCGCCGATGGC 366  
DB 102 TyrSerIleTrpGlnSerGlyGlySerLeuTyrMetGlnIleIleAsnAspLeuGly 121  
QY 367 GATCTGATCCCGCGCGCGCGCGCGCAAGCTCAAGCTTACCCATCGAGCGCAGCCGTTGCGG 426  
DB 122 GluThrIlePheThrArgValGlyLeuIleProThrIleValGlnArgTrpValPheGly 141  
QY 427 GAGGAGACGCGCAGCATCTTCGCGTGCAGATACCGCATCGGCGGCTTCGCGGCGCTC 486  
DB 142 GlnGlyAspGlySerIleLeuCyValMetAspThrGlnIleGlyArgValGlyAlaMet 161  
QY 487 TGTTGCTGGAGACATCCAGCCATTGTCAAAATACCCCATGTACGCGCGCAGCAAG 546

DB 162 CysCysTrpGlnIleIleSerGlnProLeuSerYleTyrAlaMetTyrSerGlnAspGln 181  
QY 547 GTCCACGTCGCGCTGTCGCGCGGACCTTACGCTTATTCGCGGCAATGCTTATGCTCGGA 606  
DB 182 IleHisIleAlaSerTrpProSerPheSerLeuTyrArgGlyAlaIleTyrAlaLeuGly 201  
QY 607 CCGAGGTCATTCGCGCGCGCAGCATCTACGCGGTGAGGCGCGGCTGCTACGTGCTG 666  
DB 202 ProGlnLeuAsnAsnAlaIleSerGlnMetTyrAlaIleGlnGlnCysPheValLeu 221  
QY 667 GCGTCGTCGCGCAGCGCTTTCGCGGAGATGATCAAGTATTTGATGATACCGCGCAG 726  
DB 222 AlaProCysAlaThrValSerIleGlnMetIleGlnMetLeuIleAspProArgGly 241  
QY 727 GAATGTTCTTCAAGCGCGCGCGCTTTTCCATGATTTTCGCGCGCGCGCGCGG 786  
DB 242 GluProLeuLeuLeuGlnGlyGlyGlyPheThrMetIleTyrGlyProAspGlyArgPro 261  
QY 787 CTGCGCGAGCGCGCTCCCGGAGACCGAAGGGACTGCTGCTGCGCGCATTCGACTCGG 846  
DB 262 LeuAlaIleProLeuProGlnAsnGlnGlnGlyLeuLeuTyrAlaAspIleAspLeuGly 281  
QY 847 ATGATCGCTTGGCCAAAGCGCGCGCGCATTCGCGCGCGCATTTTACGCGCGCGA 906  
DB 282 MetIleSerMetAlaIleAlaIleAlaIleAspProAlaGlyHisIleTyrAlaArgProAspVal 301  
QY 907 ACGCGGCTGCTGCTGATTCGATTCGCGCGCGCAAGCGTCTGTCACGCTTATGCGCGATT 966  
DB 302 ThrArgLeuLeuPheAsnSerAlaProAlaAsnArgValGlnTyrIleAsnProAlaSer 321  
QY 967 GAACCCCAAAACGAGACAAAGGCGAGCGCGCGCGCGCGCTGCGCGCGGGAAGGCC 1026  
DB 322 GlyPro--ThrGlnSerLeuIleAspMetGlyLeuMetGlnMetGlnAlaGlnGln 340  
QY 1027 GCCGCGCGG 1035  
DB 341 LysAlaAla 343

Search completed: April 27, 2006, 01:18:36  
Job time : 181.007 sec

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - protein search, using frame\_plus\_nzp model

Run on: April 27, 2006, 01:35:46 ; Search time 7.2946 Seconds

(without alignments)

2359.702 Million cell updates/sec

Title: US-09-751-299-1

Perfect score: 1881

Sequence: 1 atctcgagagccatgacga.....gcgcgcgcgcgcgcagtag 1041

Scoring table:

BLOSUM62	
Xgapop 10.0 , Xgapext 0.5	
Ygapop 10.0 , Ygapext 0.5	
Fgapop 6.0 , Fgapext 7.0	
Delop 6.0 , Delext 7.0	

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 1144120

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=frame+npz.model -DEV=xlp  
-Q=/abs/ABSWEB/spool/US09751299/runat\_26042006\_090039\_18719/app\_query.fasta\_1  
-DB=Issued\_Patents\_AA -OPMT=fastan -SUFFIX=rai -MINMATCH=0.1 -LOOPCL=0  
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdt  
-LIST=45 -DOCLIGN=200 -THR.SCORB=pct -THR.MAX=100 -THR.MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=pct -NOM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000  
-HOST=db8602p -USBR=US09751299\_@CGN\_1\_101\_@runat\_26042006\_090039\_18719  
-NCPU=6 -ICPU=3 -NO MMAP -NEG SCORES=0 -WAIT -DSPELOCK=100 -LONGLOG  
-DEV TIMEOUT=120 -MAIN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued\_Patents\_AA.\*

- 1: /cgn2\_6/ptodata/1/1aa/5/COMB.pep.\*
- 2: /cgn2\_6/ptodata/1/1aa/6/COMB.pep.\*
- 3: /cgn2\_6/ptodata/1/1aa/H/COMB.pep.\*
- 4: /cgn2\_6/ptodata/1/1aa/PC/COMB.pep.\*
- 5: /cgn2\_6/ptodata/1/1aa/RE/COMB.pep.\*
- 6: /cgn2\_6/ptodata/1/1aa/backfile1.pep.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	940	50.0	344	1	US-08-690-493-1
2	869	46.2	356	2	US-09-806-876A-2
3	807	42.9	369	2	US-09-823-373-5
4	807	42.9	369	2	US-09-823-373-14
5	592.5	31.5	354	1	US-08-447-702-5
6	592.5	31.5	354	1	US-08-465-615-5
7	255.5	13.6	863	2	US-08-252-991A-26099
8	247.5	13.2	467	2	US-09-252-991A-18296
9	244.5	13.0	419	2	US-09-252-991A-31734
10	242.5	12.9	724	2	US-09-252-991A-31715
11	242.5	12.9	757	2	US-09-252-991A-23569
12	239	12.7	369	2	US-09-252-991A-25394

13	238.5	12.7	344	2	US-09-252-991A-22486	Sequence 22486, A
14	238.5	12.7	681	2	US-09-252-991A-24567	Sequence 24567, A
15	237	12.6	977	2	US-09-252-991A-16655	Sequence 16655, A
16	236.5	12.6	539	2	US-09-252-991A-17070	Sequence 17070, A
17	236.5	12.6	618	2	US-09-252-991A-28358	Sequence 28358, A
18	236.5	12.6	759	2	US-09-252-991A-29395	Sequence 29395, A
19	235.5	12.5	1008	2	US-09-252-991A-29419	Sequence 29419, A
20	235	12.5	618	2	US-09-252-991A-23696	Sequence 23696, A
21	235	12.5	720	2	US-09-252-991A-21881	Sequence 21881, A
22	233.5	12.4	882	2	US-09-413-814-78	Sequence 78, Appl
23	232.5	12.4	335	2	US-09-252-991A-23674	Sequence 23674, A
24	232.5	12.4	866	2	US-09-252-991A-26065	Sequence 26065, A
25	231	12.3	552	2	US-09-252-991A-18137	Sequence 18137, A
26	229	12.2	757	2	US-09-252-991A-23068	Sequence 23068, A
27	229	12.2	681	2	US-09-252-991A-23569	Sequence 23569, A
28	227.5	12.1	611	2	US-09-252-991A-17597	Sequence 17597, A
29	227.5	12.1	722	2	US-09-252-991A-24102	Sequence 24102, A
30	226.5	12.0	511	2	US-09-252-991A-26078	Sequence 26078, A
31	226.5	12.0	541	2	US-09-252-991A-17206	Sequence 17206, A
32	225.5	12.0	545	2	US-09-252-991A-23981	Sequence 23981, A
33	225.5	12.0	438	2	US-09-252-991A-32407	Sequence 32407, A
34	225.5	12.0	1706	2	US-09-252-991A-31760	Sequence 31760, A
35	225	12.0	439	2	US-09-252-991A-17127	Sequence 17127, A
36	225	12.0	631	2	US-09-252-991A-30909	Sequence 30909, A
37	223.5	11.9	257	2	US-09-252-991A-31869	Sequence 31869, A
38	223.5	11.9	421	2	US-09-252-991A-22123	Sequence 22123, A
39	223.5	11.9	518	2	US-09-252-991A-23604	Sequence 23604, A
40	223.5	11.9	561	2	US-09-252-991A-23080	Sequence 23080, A
41	223.5	11.9	1213	2	US-09-413-814-79	Sequence 79, Appl
42	223	11.9	720	2	US-09-252-991A-21881	Sequence 21881, A
43	222.5	11.8	366	2	US-09-252-991A-22528	Sequence 22528, A
44	222.5	11.8	461	2	US-09-252-991A-31758	Sequence 31758, A
45	222.5	11.8	718	2	US-09-252-991A-32743	Sequence 32743, A

#### ALIGNMENTS

RESULT 1  
US-08-690-493-1  
Sequence 1, Application US/0860493  
Patent No. 5872000  
GENERAL INFORMATION:  
APPLICANT: Yu, Fujio  
TITLE OF INVENTION: No. 5872000el Nitrlase Gene  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Steinberg,  
STREET: 1140 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: MS-DOS Editor  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/690,493  
FILING DATE: 31 JUL-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 213061/1995  
FILING DATE: 31-JUL-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Davidson, Clifford M  
REGISTRATION NUMBER: 32,728  
REFERENCE/DOCKET NUMBER: 3821005  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 768-3800  
TELEFAX: (212) 382-2124  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 344 amino acid residues



Oy		385	CGAAGCTCAAGCCGTCACCCATGGCGGAGCGAACCCGTTCGGCGAGGGAGACGGCAGCCATC	444
Db		128	AtgylvleulysPProthrhielvalgluylgthrValPheglygluylgylturalaArgAsp	147
Oy		445	CTCGGAGTGCAGCATACCAGCCGATCCGGGCGGCGCTCGGCGGCGCTGTGTGTGGAGCACATC	504
Db		148	LeuileValserAepThrGluleuclylrgValaglYalaleucycyctPrblunhiIleu	167
Oy		505	CAGCCATTGTGAAATATACGCCATGTATAGCGCGCCGCGAACAAGTCSACAGTCGCGTGCG	564
Db		168	SerProleuserylstyYralaleuylzSerGlnhisGlualllehislellelaIatrr	187
Oy		565	CGAGGTTACAGCTCTATCGGCGCATAGGCGCTATAGGCGTCGGAACGGAGGTCAATACGGCG	624
Db		188	ProserPheiserLelytyYralaleuylzSerGlnhisGlualllehislellelaIatrr	207
Oy		625	GCAAGCCAGATCTATACGCGATCGAGGCGCGACTGCTACAGTGCCTGCGCTGCGCGACGCTT	684
Db		208	AlaserGlnilietrySerValgluclYgelncysPheThrillelaIalaserSerValVal	227
Oy		685	TGCGCCGAGATGATCAAGGTATGTGGATACGCCCCGACMAAGAGATGTTCTCAAGGCC	744
Db		228	ThringlutrlrLieuAspMetleuclYvalglYgluHisAmAlaProleuLieuylsVal	247
Oy		745	GGCGGCGGTTTTGGCATGATTTTGGGGCGCGAGCGGCGCGGCGCTGGCGAGCGCGTCCG	804
Db		248	GlYglYglYserSerMetellePheIalProaspyglayagThrLieuAlaProTyLieuPro	267
Oy		805	GAGACCGAAGAGGAGCATCTGGTCCGCCGATATGCACTCGGCATGATGCGTTGGCGCAAG	864
Db		268	HIsapRlalagluYleuilellelaIaspyeuanMetgluclnullelaIaPhehalays	287
Oy		865	GCGCGCGCGGATCCGGCGGCGCATATTCACGGCCCCGACGTAACGCGGCTGCTGGCAT	924
Db		288	AlaIleaaenApPrCoValglYHietrySerLysProGluallathrArgLieuValLieuAsp	307
Oy		925	-----CGACGTCGGCGCCCAAGCGGCGTCGACAGCTTGATCGCGCATTGGAACCGGAA	975
Db		308	LenglYHisAlaArgAepPrometThrArgVal-----HiserylS	320
Oy		976	AACGAGACAAGGCGGACGCGCGCGCTGCGGTAGCGGAAAGCGCGCGCGCGCG	1035
Db		321	SerValThrArggluGlualaProdlugluInglYValGlnserLytlelaIaserValala	340
RESULT 3				
US-09-823-373-5				
; Sequence 5, Application US/09823373				
; Patent No. 6870038				
GENERAL INFORMATION:				
APPLICANT: Chauhan, Sarita				
APPLICANT: Dicostimo, Robert				
APPLICANT: Payne, Mark				
APPLICANT: Gavagan, John				
TITLE OF INVENTION: Isolation and Expression of a Gene for Nitritease from				
FILE REFERENCE: BC-1032 US NA				
CURRENT FILING DATE: 2001-03-30				
PRIOR APPLICATION NUMBER: 60/1193,707				
NUMBER OF SEQ ID NOS: 32				
SOFTWARE: Microsoft Office 97				
SEQ ID NO 5				
LENGTH: 369				
TYPE: PRT				
ORGANISM: Acidovorax facilis				
US-09-823-373-5				
Alignment Scores:				
Pred. No.:	1.18e-61	Length:	369	
Score:	807.00	Matches:	160	

[illegible]

Db 325 AepValAsnThrArgSerArgValGIuAsnPheArgLeuArgGlnIalalagIuGlnIu 344  
Qy 1027 GCCGCCGCCG 1035  
Db 345 ArgGlnAla 347

RESULT 4  
 US-09-823-373-14  
 : Sequence 14, Application US/09823373  
 : Patent No.6670038  
 : GENERAL INFORMATION:  
 : APPLICANT: Chauban, Sarita  
 : APPLICANT: DiCosimo, Robert  
 : APPLICANT: Payne, Mark  
 : APPLICANT: Gavagan, John  
 : APPLICANT: Fallon, Robert  
 : TITLE OF INVENTION: Isolation and Expression of a Gene for Nitriase from  
 : TITLE OF INVENTION: Acidovorax Facilis 72W  
 : FILE REFERENCE: BC-1032 US NA  
 : CURRENT APPLICATION NUMBER: US/09/823,373  
 : CURRENT FILING DATE: 2001-03-30  
 : PRIOR APPLICATION NUMBER: 60/193,707  
 : PRIOR FILING DATE: 2000-03-31  
 : NUMBER OF SEQ ID NOS: 32  
 : SOFTWARE: Microsoft Office 97  
 : SEQ ID NO 14  
 : LENGTH: 369  
 : TYPE: PRT  
 : ORGANISM: Acidovorax facilis  
 : US-09-823-373-14

	US-09-751-299-1 (1-1041)	x	US-09-823-373-14 (1-369)
Alignment Scores:			
Pred. No.:	1,18e+61		Length: 369
Score:	807.00		Matches: 160
Percent Similarity:	61.8%		Conservative: 52
Best Local Similarity:	46.6%		Mismatches: 127
Query Match:	42.9%		Indels: 4
DB:	2		Gaps: 2

APPLICATION NUMBER: US 08/194,588  
FILING DATE: 10-FEB-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: FR 92-09-882  
FILING DATE: 10-AUG-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: McGowan, Malcolm K.  
REGISTRATION NUMBER: 39,300  
REFERENCE/DOCKET NUMBER: 003025-019  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 703-836-6620  
TELEFAX: 703-836-2021  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 354 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-447-702-5

Alignment Scores:  
Pred. No.: 4.76e-43 Length: 354  
Score: 592.50 Matches: 137  
Percent Similarity: 54.0% Conservative: 51  
Best Local Similarity: 39.4% Mismatches: 133  
Query Match: 31.5% Indels: 27  
DB: 1 Gaps: 7

US-09-751-299-1 (1-1041) x US-08-447-702-5 (1-354)

31 GCGGCGGTGACGCGCGCGCGCTTCCTGATTCGACCGACGATCGAAGAGCATC 90  
10 AAlaAlaValGlnAlaAlaProValPheMetAsnLeuGlnAlaTrpValAspLysThrCys 29  
91 GGCCTGATGACGACGCGCGCGACGACGACGCGCTGATTCGATTCGACGATTCG 150  
30 LysLeuIleAlaGlnAlaAlaSerMetGlyAlaLysValIleGlyPheProGlnAlaPhe 49  
151 ATTCCCGGCTATCCCTTTTGATATGGCTGGC-----GCGCGGCTTGG 195  
50 IleProGlyTrpProGlyTrpIleTrpThrSerAsnMetAspPheTrpGlyMetMetTrp 69  
196 GGCATGGCTTCGTCGACGCGCTATTCGAGATTCGCTCGCGCGCGACGACGACGTG 255  
70 AlaVal-----LeuPheLysAsnAlaIleGluIleProSerLysGlnVal 84  
256 CAGGCGCTGCGGATGCGCGCGCGCGCGACGACGACGATGCTGCGCGCTATGCGAG 315  
85 GlnGlnIleSerAsnAlaAlaLysLysAsnGlyValTrpValCysValSerValSerGlu 104  
316 GCGCGCGGCGCGACGCTCTATATGGCGACGCGATTCGCGCGCGATGCGATTCATC 375  
105 LysAsnAsnAlaSerLeuTrpLeuThrGlnLeuTrpPheAspProAsnGlyAsnLeuIle 124  
376 GCGCGCGCGCGACGACGACGACGACGACGACGACGACGACGACGACGACGACGAC 435  
125 GlyLysHisAlaGlyLysPheLysProThrSerSerGlnTrpAlaValTrpGlyAspGly 144  
436 GCGACGATTCGCGGCTGCGACGATACCGCATCGCGCGCTGCGCGCTGCTGTTGG 495  
145 GlySerMetAlaProValPheLysThrGluTrpGlyAsnLeuGlyLysLeuGlnCysTrp 164  
496 GAGCATTCACGACGATTCGAAATACGCAATGTAACGCGCGCGACGACGATTCACGTC 555  
165 GlnHisAlaLeuProLeuAsnAlaAlaIleMetCysLeuAsnGlnGlnValHisVal 184  
556 GCGGCGGCGCGCGCTC-----AGCCTCTATCGCGGATGCGC 594  
185 AlaSerTrpProAlaPheValProLysGlyAlaValSerSerTrpValSerSerVal 204  
595 TATGCGCTCGACGACGATCAATACCGCGCGACGACGACGATTCACGCGCTCGAGCGCGC 654  
205 CysAlaSerThrAsnAlaMetHisGlnIleIleSerGlnPheTrpAlaIleSerAsnGln 224

655 TGCTAGTGTGGCGGTGTCGCGGACCGCTTCGCGCGAGATGATCAAGTATTGGTGGAT 714  
225 ValTrpAlaIleMetSerThrAsnLeuValGlyGlnAspMetIleAspMetIleGlyLys 244  
715 ACGCCGACAGAGATGTTCTCAAGCGCGCGCGCTTTCGATTCGATTCGAGGCC 774  
245 AspGlnPheSerLysAsnPheLeuProLeuGlySerGlyAsnThrAlaIleIleSerAsn 264  
775 GAGGCGCGCGCGCTGCGCGCGCGCTCCGGAACGACGACGACGATTCGTCGCCGAT 834  
265 ThrGlyGluIleLeuAlaSer---IleProGlnAspAlaGluGlyIleAlaValAlaGlu 283  
835 ATGACCTCGGACGATGATCGGCTTGCCCAAGCGCGCGCGATTCGCGCGCGCTATTC 894  
284 IleAspLeuAsnGlnIleIleTrpGlyLysTrpLeuLeuAspProAlaGlyHisTrpSer 303  
895 CGCGCGACGTTAAGCGCGCTGCTGCTGATCGACGT-----CGCGCCCAACGCGTC 945  
304 ThrProGlyPheLeuSerLeuThrPheAspGlnSerGlnHisValProValLysGlyIle 323  
946 -----GTACGCTTGATGCGCGATTCGAAACGCAAAACGACGACGACGAC--- 993  
324 GlyGlnGlnThrAsnHisPheIleSerTrpGluAspLeuHisGlnAspLysMetAspMet 343  
994 -----GCGCGCGCGCTGCGCGCTG 1011  
344 LeuThrIleProProAlaGlyVal 351

RESULT 6  
US-08-465-615-5  
Sequence 5, Application US/08465615  
Patent No. 5635391  
GENERAL INFORMATION:  
APPLICANT: PETRE, Dominique  
APPLICANT: CERBELEAUD, Edith  
APPLICANT: LEVY-SCHIL, Sophie  
APPLICANT: CROUZET, Joel  
TITLE OF INVENTION: POLYPEPTIDES POSSESSING A NITRILASE  
TITLE OF INVENTION: ACTIVITY, DNA SEQUENCE CODING FOR SAID POLYPEPTIDES,  
TITLE OF INVENTION: EXPRESSION CASSETTES AND HOST MICROORGANISMS ENABLING THEM  
TITLE OF INVENTION: TO BE OBTAINED, AND METHOD OF CONVERTING NITRILES TO  
TITLE OF INVENTION: CARBOXYLATES BY MEANS OF SAID POLYPEPTIDE  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESSES:  
ADDRESSER: Burns, Doane, Swecker & Mathis  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/465,615  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/194,588  
FILING DATE: 10-FEB-1994  
APPLICATION NUMBER: FR 9209882  
FILING DATE: 10-AUG-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Crane-Feury, Sharon E  
REGISTRATION NUMBER: 36,113  
REFERENCE/DOCKET NUMBER: 003025-015  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 5:



```

Db      338  |||
          GlyThrIaIaIaAaGArGThrHiAaRProTrpProGlyAaGlyAlaIaGlyIaAspAla 357
Qy      351  ---CTTCGGCCC-----CGATGGCATGTATCGCGCGCCGCGCA 389
Db      358  |||
          LeuLeuArGProAaPAlaIaGlyIaIaIaAaPArGLeuArGlyAaPProArGArSerPro 377
Qy      390  GCTCAAGCTTACCA---TGGGAGCGACCGTGTGGCGAGGAGACCGCACCATCT 446
Db      378  |||
          AlaGlnIaAaPProAaGAlaIaGlyIaIaIaAaRAlaIaIaGlyIaAaRgGlnProAla 397
Qy      447  -----CGCGGTGACGATCCGCGATCGGCGC 473
Db      398  ValProArGlnProProGlyAlaSerGlyLeuArGArGArGlyAaPArGLeu 417
Qy      474  CTTGCGCGCGCTGTGTGCTGGAGACATCCAGCATTTGCAAAATACCGCATGTACG 533
Db      418  |||
          HisArGArGThrLeuArGArGArGProGlnProAlaThrAlaPro-----AlaGly 434
Qy      534  CGCCGACGAAGATCCACGTCGCGCTGGCCGACGCTTACGCTTATCGCGCATGCG 593
Db      435  ArGArGlnGlyThrGlyAaRgArGProAla-----LeuProGlyAaRgArGly 452
Qy      594  CTATGCGCT-----CGGACCGGAGTCAATACCGCGC 626
Db      453  GlnCyAlaPProAlaArGProGlyAaRgLeuAlaIaIaArGArGlyThrValGlnAaP 472
Qy      627  AAGCAGATCTACCGCGTCGAGGCGCGCTGCTACGTGCGCTGCGCGACCGTTTC 686
Db      473  |||
          ArgProGlnIaAaPArGArGThrAlaArGlyAaPArGProGlyAaRgProArGlyAaRgSer 492
Qy      687  GCCGGAATATCAAGTATTGTGTGATACCCCGACAGAGATGTTCTTCAAGCGCG 746
Db      493  ProAlaIaThrAlaIaGlyAlaPProGlyAlaThrArGArGArGAlaPProAlaArGLeu 512
Qy      747  CGGCGGTTTTCATGATTTTGGGCGCGACCGCGCGCT----- 788
Db      513  ArGArGThrLeuGlyIaIaArGArGThrArGArGHisSerProProArGAlaPProArG 532
Qy      789  GAGCGGACCGCTCC--GAGACCGA-----AGAGGACT 821
Db      533  GLyArGArGAlaPProLeuGlyIleArGProArGArGlnProArGThnGlyArGlyHis 552
Qy      822  GCTGCTGCCGATTCGACT----- 842
Db      553  ProArGArGArGlnArGProGlyAaRgArGlnPheArGHisAlaIaGlnGlyAlaAaPgly 572
Qy      843  ---CGGATGATGCGGTTGGCGAAGGCGCGCGCATCCGCGGCGACATTTACCGGCC 899
Db      573  ArGArGHisArGArG-----GlyAlaGlyArGArGlyArG-----ThrGly 586
Qy      900  CGACGTAAACGCGCTGCTGTGATCG-----ACGTCGCGC 935
Db      587  ArgGlnGlyProThrAlaIaIaGlyAlaGlyIaIaAaPArGLeuArGlyThrProAlaGlnGly 606
Qy      936  CCAACGCTGCTCAACGCTTGATCCGCAATTGAAACCGAAACGAGACAGGACGCGC 995
Db      607  ArGArGArGArGArGlyProProArGProGlyAaRgProTrpArGAlaIaGlnPro 626
Qy      996  GCCCGCGTGGCGTGGTGGCGGAAAGCGCGCGC 1031
Db      627  AlaGlyAlaGlyArGlyGlySerArGArGLeuArG 638

```

```

RESULT 8
US-09-252-991A-18296
; Sequence 18296, Application US/09252991A
; Patent No. 6551795
; GENERAL INFORMATION:
; APPLICANT: Marc J. Rubenfield et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
; FILE REFERENCE: AERGINOSA FOR DIAGNOSTICS AND THERAPEUTICS
; CURRENT APPLICATION NUMBER: US/09/252,991A

```

```

; CURRENT FILING DATE: 1999-02-18
; PRIOR APPLICATION NUMBER: US 60/074,788
; PRIOR FILING DATE: 1998-02-18
; PRIOR APPLICATION NUMBER: US 60/094,190
; PRIOR FILING DATE: 1998-07-27
; NUMBER OF SEQ ID NOS: 33142
; SEQ ID NO 18296
; LENGTH: 467
; TYPE: PRT
; ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-18296

Alignment Scores:
Pred. No.: 4,57e-13 Length: 467
Score: 247.50 Matches: 119
Percent Similarity: 34.3% Conservative: 15
Best Local Similarity: 30.4% Mismatches: 126
Query Match: 13.2% Indels: 131
DB: Gaps: 21

US-09-751-299-1 (1-1041) x US-09-252-991A-18296 (1-467)
Qy      1033  CGGCGCGCGCGCTTTCGCCACGACGCGCGCGCGCGCTGC----- 989
Db      106  ArGArGArGArGHisPProAlaIaTrpArGArGArGArGArGProAlaIaGlyAlaPro 125
Qy      988  -----CTTGTCTCGTTTGGCGTTGCAATGCGGCATCAAGCGTGAACGCGCTT 938
Db      126  ProProAlaProCyPro-----AlaProArGlyAla 136
Qy      937  GGGCGGAGCTCATCCAGACGAGCGCGCTTACGTGCGGCGGTAATAGTGGCGCGCG 878
Db      137  GLy-----AlaProSerValArGlnGlyProAlaIaGlySerAlaIaProPro 151
Qy      877  GATCGGCGC-----CCGCTTGGCCAACG-----CGA 851
Db      152  ProArGProAlaArGArGProProAlaProProHisArGThrGlySerGlyTrpArG 171
Qy      850  TCATGCCAGGTGATTCGCGACGACGACGATCCCTTGTGTTCCGGA-----GCG 797
Db      172  ArGArGArGlyArGThrArGArGProAlaIaIaSerArGArGProAlaIaGlyIaProAla 191
Qy      796  GCTCGCGAGGCGCGCGCGCTGCGGCCGCAAAATCATGCAAAACGCGCGCGCTTGA 737
Db      192  GLyArGProGlyIaGlySerArG-----AaArGArGArG----- 202
Qy      736  GGAATCTCTCTGTGCGGCGGTATCCACATATCTTATCATCTCGCGGAAACGCTGC 677
Db      203  ---AlaGlyCyArGProAlaArGProAla----- 211
Qy      676  CGACGACGCGCACGATGACGCGCGCTCGACCGCGGTAGATCTGGCTTGGCGGCTAT 617
Db      212  -----AlaArGArGAlaIaGlyProGlyArGCyArGArGArGArGArGArG 227
Qy      616  TGAATCCGCTCCGAGCGCATAGGCATGC-----CGCATAGAGGCTGAAGCTCG 566
Db      228  -----AlaTrpSerProCyProThrProArGArGArG-----AlaGly 240
Qy      565  GCCACGACGCGACGATGACCTGTTCGTCGCGGCGGTACATGAGGTATTTGACAAATGCT 506
Db      241  AlaGlyAaPArGHisValPro-----Ala 248
Qy      505  GATGTGCTCCCAACGACAGAGCGCGCGAGCGCGCGATGCGGTATCGTGCACGCGGA 446
Db      249  GLyAaPAlaArGAlaIaIaGlyTrpProThnGlySerArGyAlaIaRgArGSerGlySer 268
Qy      445  GATGCTGCGCGCTCTCTCGCGACACAGCGGTGCTCCGATGGTGAAGCTTACGCTGC 366
Db      269  AlaGlyCyAlaArGProProArGArGArGArGArGArGArGArGArGArGArGArG 282
Qy      385  GGGCGCGCGGATCGCATCGCATCGGCGCGGAGATCGCTGCGCATATAGAGGCTGC 326
Db      283  AlaIaIaArGArGValaIaArGArGSerAlaArGArGSerArGProValaArGSerAla 302

```

```
QY 325 -----CGCCGCGCGCTTCTGCTATAGCCGCCACAGATGCGCTGGCGCGCG 275
Db 303 LeuArgGlyLeuProAlaAlaArgSerArgProArgHis-----ArgArgProGly 320
QY 274 CCGGATCGCGCGGCGCTGACACTGCTTGCCTGCGCGCAGAGGAATTCTCGAAATAGC 215
Db 321 SerValProPro-----ValArgThrSerArgAlaThrProAlaArgArgTrpHisAla 337
QY 214 GCTGACGAGACGCGATGCCACCGCGCGCGCCAGCCATATCCAAAAGGATAGCCGG 155
Db 338 ArgTrpGlnAlaAlaLeuArgArgProGly----- 347
QY 154 GAATCCAAAGTCTGGGATGCGATCAGCGCAGCTCTGCTGGCGCGCTGCTGATCA 95
Db 348 -----SerValLeuSerSerAlaThrProAlaLeuProGlyValArgGly 362
QY 94 ---GGCCGATCGCTTTCTGA-----CTGTGCGGTGAGATCGAGACA 53
Db 363 TyrProArgGlyTyrAlaArgAlaProArgProAlaLeuProGlyArgProAlaArg 382
QY 52 CCGGCGGCGCTGACCGCGCGCGCGGAT----- 23
Db 383 ProAlaArgProArgProCysArgArgProAlaAspArgSerProGlyGlnGlyArgPro 402
QY 22 -----ACTTCGTGATGGGCTCCG 5
Db 403 ProGlySerSerAlaThrArgArgTrpArgPro 413

RESULT 9
US-09-252-991A-31734
/ Sequence 31734, Application US/09252991A
/ Patent No. 6551795
/ GENERAL INFORMATION:
/ APPLICANT: Marc J. Rubenfield et al.
/ TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
/ FILE REFERENCE: 107196.136
/ CURRENT APPLICATION NUMBER: US/09/252,991A
/ CURRENT FILING DATE: 1999-02-18
/ PRIOR APPLICATION NUMBER: US 60/074,788
/ PRIOR FILING DATE: 1998-02-18
/ PRIOR APPLICATION NUMBER: US 60/094,190
/ PRIOR FILING DATE: 1998-07-27
/ NUMBER OF SEQ ID NOS: 33142
/ SEQ ID NO 31734
/ LENGTH: 419
/ TYPE: PR
/ ORGANISM: Pseudomonas aeruginosa
US-09-252-991A-31734

Alignment Scores:
Pred. No.: 7,98e-13 Length: 419
Score: 244.50 Matches: 128
Percent Similarity: 37.7% Conservative: 28
Best Local Similarity: 30.9% Mismatches: 134
Query Match: 13.0% Indels: 124
DB: Gaps: 23

US-09-751-299-1 (1-1041) x US-09-252-991A-31734 (1-419)
QY 24 TCGCCGCGCGCGC-----GCTGCAGGCGCGCGGCTGTTCTTCTGA 62
Db 6 AlaArgArgGlyGlyArgArgAspAlaProHisGlyAlaGlyHisArgAlaArgArgPro 25
QY 63 TCTGCAACGACAGTGGAAAGGATGGGCTGATGAGACAGCGCGCCAAAGACAGACGT 122
Db 26 AlaArgProHis--ArgProAlaArgArgAlaAspArgProGlyHisArgSerAlaArg 44
QY 123 GCGCCTGATGCAATCCGACAGACTTGATCCCGGCTATCCCTTTGATATGCTGGG 182
Db 45 ArgProAlaAlaArgArgArg-----SerGly 53
```

```
QY 183 CCGCGCGCGCTTGGGAGATGCGCTTCTGACGCGCTATTTGAGAAATTGCTGCGCGG 242
Db 54 TrpSerGlyAlaArgLeuAlaArgGlyGlyCysAlaArgArg-----CysArgProAsn 71
QY 243 CAGAGACAGATGGACAGGCGCTTGGCGATGCGGCGCGCGCGCGGACGATGATCGCGG 302
Db 72 ArgArgHisValProGlyThrProGlyAlaGlyTyrArgProArgArgThrCysSerAla 91
QY 303 CCGCTATAGGACGCGC-----GGGCGGACGCTTATATGGG----- 341
Db 92 ProArgAsnProProAlaArgArgTrpProGlyGlyArgArgProThrSerGlyAsnAla 111
QY 342 ---CGAAGCTATGTCGCGCGCGCGATGCGATCTGATGCGCGCGCGCGGACGTCGAAGC 398
Db 112 ArgProGlyProValAlaArgAlaArgTrpArgGlnProGlyCysSerPro----- 127
QY 399 TACCGATGCGAGACGACCGCTGTTGCGGAGGAGACGCGACGATCTCGGTCGACGA 458
Db 128 -----GlyGlyHisArgGlyArgGlnHisAlaGlyTrp-----SerArgAspArgArg 142
QY 459 TACCGCATGCGGCGCGCTGCGCGCGCTCTGCTGCGAGACATCC-----A 506
Db 143 His--ProArgGlyArgSerValArgGlnArgArgCysSerAspSerValGlyAspAsp 162
QY 507 GCCATTGTCGAAATACGCGCATGTACGCGCGCGGACGAAACAGTCCAGTCCGTCGCGC 566
Db 162 rArgSerArgSerLeuProProAspGlyProAlaGlyArgSerGlyThrAlaProGlyA 182
QY 567 GAGCTTCAGCTCTATCGCGGATGGCT----- 595
Db 182 rArgSer-----ArgTrpSerGlyArgProAlaCysArgLeuArgArgAlaG 198
QY 596 -----ATGCGCTCGACCGGACGATCAATACCGCGCGGACGATCTTACGC 641
Db 198 LysArgThrLeuArgArgArgArgArgGlySerProAlaProAlaAspArgArgAlaPro 218
QY 642 GGTGCAAGGCGGCTGCTAGCTGCTGCGCTGCGCGGACCGCTTCCCGGAGATGACAA 701
Db 218 rGlyS-----ThrAlaTrpArgAlaProArgHis----- 227
QY 702 GGATATTGATGATACCGCGACGAGAGATGTTCCCAAGCGCGCGCGGTTTGGCAT 761
Db 228 -----Arg-ArgTrpGlySerAlaAlaGlnGlyArgArg-----IleHis 240
QY 762 GATTTT---CGGCGCGGACGCGCGCGCTGCGCGGAC---GCTCCGAGACGACGA 815
Db 241 SerSerGlyArgCysArgArgProIleProGlyArgSerValAlaProAlaGlyArgSer 260
QY 816 GGGAAT-----GCTGCTCGCGCATATGCACT----- 842
Db 261 GlyAlaGlySerValProAlaGlyArgCysTrpArgProAsnValLeuArgCysArg 280
QY 843 -----CGCATGATCGCGTTGSCCAAGCGCGCGGATCTCGCG 881
Db 281 SerArgAsnAlaArgProGlyArgAsnAsnArgSerThrArgProThrGlyArgSerGly 300
QY 882 GGGCAGCATATTCACGCGCGCGACGTAAACGG-----GCTGCTGTCGGA---TGG 926
Db 301 SerThrProGlyAlaAlaAlaArgArgArgAlaGlyArgArgPheAlaSerAlaGlyArgAla 320
QY 927 ACGTCCGCGCCAAACGCTGCTACGCTTATGCGCG----- 962
Db 321 GlyAlaGlyLeuAlaArgArgArgSerGlyCysValArgProGlyArgAlaProSerSerHis 340
QY 963 -----ATTGCAACGCGCAAAAGAGAGCAAGGCGGACGCGCGCG----- 1001
Db 341 CysProValAlaArgArgAlaArgArgSerSerGlyAlaAlaSerArgProProAlaArgPro 360
QY 1002 GCTGCGCGCTGATGCGGAAAGCGCGCGCGCGCGCA 1037
Db 361 AlaAlaArg-----SerLeuArgArgArgAla 369

RESULT 10
```









QY 488 CAGAGCGCGCGAGCGCGCGATATCGTCAACCGAGATGGCTGCTCC 429  
Db 153 -----ArgTTPSerThAlaArgThRlCysSerValPr 164  
QY 428 T-----CGCCGAACACGGTGCCTCGCATGGGTAGGC 396  
Db 164 oSerGlyAlaLeuArgProTTPSerGlyArgAspArgArgCysThRlGlyAlaGlySerAl 184  
QY 395 TTGAGCTTGGCGCGCGCGCATCAGATCCGCATCGGCGC-----CG 354  
Db 184 aAlaAlaValArgGlyArgArg---AspGlyTTPArgGlyAlaGlyAsnArgAspTTPAr 203  
QY 353 AGATGCGCTTGGCGCGCATATAGAGCTGCCCGCGCGCTCGCATATGCCGCGCACAGACA 294  
Db 203 gGluCysProGlyPro-----GlyAlaProProArgThr-----ProLeuAr 217  
QY 293 TGCATGCCGT-----GGCGCGCGCGCGCATCCCGACGGCTTGCATCTGCTGCTG 243  
Db 217 gAlaAlaArgArgValAlaGlySerArgAspTTPProGluCysProGlyArgAla----- 235  
QY 242 CCGCGCACGAGCGCATCTTCCGAATAGCTTGACGAGACGATGCCCGCCAGCGCGCGC 183  
Db 236 ----ValArgArgArgThRArgSerAlaAspArgArgAlaAlaGlyCysArgGlyPh 254  
QY 182 CCGAGCCATATCCAAAGAGTAGCCGGGATCCAGTCTCTGGGAATGCGATCAGGCGC 123  
Db 254 eProGlyCysArgGlyArgAlaAlaArgGlu---LysThRArgGlyAlaThRArgAlaAl 273  
QY 122 ACCTCCCTCTTGG-----CGCTGCT-----CGATCAGCGCGATCGCTTTC 81  
Db 273 aSerGlyCysArgThRThRArgSerProAlaValAlaGlyArgThRArgArgGluValG 293  
QY 80 TCGACTGTGC 71  
Db 293 yPheValCys 296

RESULT 14  
US-09-252-991A-24567  
; Sequence 24567, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 24567  
; LENGTH: 681  
; TYPE: PR  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-24567

Alignment Scores:  
Pred. No.: 3,21e-12 Length: 681  
Score: 238.50 Matches: 126  
Percent Similarity: 33.9% Conservative: 19  
Best Local Similarity: 29.4% Mismatches: 152  
Query Match: 12.7% Indels: 131  
DB: 2 Gaps: 20

US-09-751-299-1 (1-1041) X US-09-252-991A-24567 (1-681)  
QY 1036 GCGCGCGCGCGCGCTTTCCGCCACCAACCGCGCGCGCGCGCTGC----- 989  
Db 79 SerLyArgSerAlaProProAlaProAlaSerThRArgThRAlaArgGly 98

QY 988 ---CCTTGT-----CCTGCTTTGGGTTTGAATCGCGCATACGC----- 950  
Db 99 SerProCysAlaTTPMetProLyAlaAlaCysArgArgArgProAlaAlaAlaMetArg 118  
QY 949 -----TGACGACGCGTGGGCGCGACGTC 926  
Db 119 SerCysArgTTPSerSerGlyAsnGlyArgTTPArgTTPArgArgArgAlaValArgPro 138  
QY 925 GATTCAGCAGACAGCCCGTTACGTCCGGCCGTGAATAGTGCCCGCGCATCGCGCGC 866  
Db 139 TTPProAlaSerProArgValAlaArgValAlaGlyValTTPProProSerLeuCysPro 158  
QY 865 CTTTGGCAGACGAGATCATGCGCGCATATCGGCGACAGCAGTCCCTTT----- 812  
Db 159 AlaProProArgArgArgArgProGlyArgGlnArgGlnArgProArgGlyProAlaAlaGln 178  
QY 811 -----CGGTCTCCGGA-----GCGGCTTCCGCGCGCGC 779  
Db 179 ArgArgAlaAlaTTPArgArgArgAlaGlyAspGlnProAlaAlaArgProGlyArgGly 198  
QY 778 CGTGCG-----GCTGCTCCGGA-----GCC 770  
Db 199 GlnArgThCysLeuGlnProLeuAlaValAlaGlnProAlaValLeuGlnCysProAla 218  
QY 769 CGAAATCATGCGCAAAACCGCGCGCGCTTGAGAACATCT----- 728  
Db 219 ArgArgAlaGlyAspHisSerThRArgGlySerGlyThCylAspProSerArgArg 238  
QY 727 CTTTGGCGCGGTATCCACCATCTTATCATCTCCGCGCAACGC----- 680  
Db 239 ProGlyArgGlyAspGlyPro---ProGlyGlnProGlyAlaAspArgLeuAspSerGly 257  
QY 679 -----TCGCGCACGACGCGCGCATGACGCGC----- 650  
Db 258 GlyGlyProAlaHisAlaProAlaAlaProThProProArgLeuHisArgAlaArgGln 277  
QY 649 -----CCTGACCGCGGTAGATCTGCTGGCGGATTTGACCTTCGCTCGAGCG 599  
Db 278 SerLeuAlaProArgProArgArgIle-----ProProArgSer 290  
QY 598 CATAG-----CCATCCCGGATAGAGCTGAAGCTCGCGCACAGC 557  
Db 291 ArgArgGlyProAlaAlaProLeuProLeuArgAsp-----Pro 303  
QY 556 CGACGTGACCTGTTGCGCGCGCGCTGACATGCGATTTTCACATAGCTGAGTGTCT 497  
Db 304 AlaGlyAlaProGlyArgArgPro-----ArgLeuGlnThRLeuAlaGlyAlaAla 320  
QY 496 CCGAGCAACAGACCGCGCGCGCGCGCGATGCGGTATCGTGCACCGCGCATGCTGC 437  
Db 321 AlaProProArgArgAlaGlyGlyArg---ValLeuArgArgSerArgArgGlySer 339  
QY 436 CGTCTCCCTCGCGCAACACGATGCGCTCCGATAGGTAGGCTTACCTTGC----- 386  
Db 340 GlyLeuProCysLeuArgProValAlaAspProThLeuProAlaThRArgCysProAlaPro 359  
QY 385 GCGCGCGCGCATCAGATCGCATCGGCGCGAAGATCGCTTGGCCCATATAGAGCTGC 326  
Db 360 GlyAlaArgGlyArgAspGlyHisArgSerGlyArgArgProGly-----GlyGly 376  
QY 325 CGCGCGCGCGCTCCGAT-----AGCGCGCACGACATGCA 290  
Db 377 ArgArgArgArgArgGlnProGlyLeuValAlaAlaGlyArgAlaAlaArgThRArgGlu 396  
QY 289 TGCCTGGGCGCGCGCGCGCATCCGCGAGGCGCTGCACTGTGTCGCGCGCACGAGCG 230  
Db 397 CysArgArgAlaAlaGlyHisProGlnProValArgThGlyLeuGlyArgThRArgVal 416  
QY 229 AATTTCGAATATAGCGCTGACGAAGCGCATGCCCAACCGCGCGCGCCACCATATCC 170  
Db 417 GlnSerArgArg-----ProAlaProProArgLeuProGlyLe 429  
QY 169 AAAAGGATAGCGCGGATTCGAAGTCTTGGAATGCGATGAGCGCAGTCTGCTGCTG 110



**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - protein search, using frame\_plus\_nzp model

Run on: April 27, 2006, 01:38:31 ; Search time 32.8257 Seconds  
(without alignments)  
2650.121 Million cell updates/sec

Title: US-09-751-299-1

Perfect score: 1881  
Sequence: 1 atctgcgagccatgacgaa.....gcgcgcgcgcgcgcagtag 1041

Scoring table:  
BIOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 3735138

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODEL=frame+npz.model -DEV=xlp  
-Q=/abs/ABSSMB.spool/US09751299/runat 26042006 090042 18780/app query.fasta\_1  
-DB=Published Applications AA Main -OPMT=fastan -SUFFIX=rapbm -MINMATCH=0.1  
-LOOPC=0 -LOOPEXT=0 -UNITS=bits -START=1 -MATRIX=blonsum62  
-TRANS=human40.cdi -LIST=45 -DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100  
-THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pco -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=2000000000 -HOST=abss802p  
-USBR=US09751299 @CGN 1.1 405 @runat 26042006 090042 18780 -NCPU=6 -ICPU=3  
-NO MMAP -NEG SCORES=0 -WALT\_DSPBLOCK=100 -LONGLOG -DEV TIMEOUT=120  
-WALT TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FAPOP=6 -FAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications AA Main:\*

- 1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep:\*
- 2: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep:\*
- 3: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep:\*
- 4: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep:\*
- 5: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep:\*
- 6: /cgn2\_6/ptodata/1/pubpaa/US11\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1806	96.0	346	3	US-09-751-299-2
2	1806	96.0	346	3	US-10-146-772-384
3	1806	96.0	346	4	US-10-241-742-16
4	1806	96.0	346	4	US-10-241-742-16
5	1806	96.0	346	4	US-10-440-523-350
6	1806	96.0	346	4	US-10-440-523-350
7	1806	96.0	346	4	US-10-461-925-16
8	1464	77.8	341	4	US-10-241-742-34
9	1464	77.8	341	4	US-10-440-523-34
10	1464	77.8	341	4	US-10-440-523-34
11	1464	77.8	341	4	US-10-461-925-34

12	1209	64.3	348	4	US-10-146-772-16	Sequence 16, App1
13	1209	64.3	348	4	US-10-241-742-16	Sequence 16, App1
14	1209	64.3	348	4	US-10-440-523-16	Sequence 16, App1
15	1209	64.3	348	4	US-10-461-925-16	Sequence 16, App1
16	1209	64.3	348	4	US-10-461-925-16	Sequence 16, App1
17	1165	61.9	353	4	US-10-146-772-212	Sequence 212, App
18	1165	61.9	353	4	US-10-241-742-212	Sequence 212, App
19	1165	61.9	353	4	US-10-440-523-212	Sequence 212, App
20	1165	61.9	353	4	US-10-440-523-212	Sequence 212, App
21	1165	61.9	353	4	US-10-461-925-212	Sequence 212, App
22	1155	61.4	353	4	US-10-146-772-232	Sequence 232, App
23	1155	61.4	353	4	US-10-241-742-232	Sequence 232, App
24	1155	61.4	353	4	US-10-440-523-232	Sequence 232, App
25	1155	61.4	353	4	US-10-440-523-232	Sequence 232, App
26	1155	61.4	353	4	US-10-461-925-232	Sequence 232, App
27	1148	61.0	353	4	US-10-146-772-350	Sequence 350, App
28	1148	61.0	353	4	US-10-241-742-350	Sequence 350, App
29	1148	61.0	353	4	US-10-440-523-350	Sequence 350, App
30	1148	61.0	353	4	US-10-440-523-350	Sequence 350, App
31	1148	61.0	353	4	US-10-461-925-350	Sequence 350, App
32	1132.5	60.2	358	4	US-10-146-772-170	Sequence 170, App
33	1132.5	60.2	358	4	US-10-241-742-170	Sequence 170, App
34	1132.5	60.2	358	4	US-10-440-523-170	Sequence 170, App
35	1132.5	60.2	358	4	US-10-440-523-170	Sequence 170, App
36	1132.5	60.2	358	4	US-10-461-925-170	Sequence 170, App
37	1130	60.1	354	4	US-10-146-772-102	Sequence 102, App
38	1130	60.1	354	4	US-10-241-742-102	Sequence 102, App
39	1130	60.1	354	4	US-10-440-523-102	Sequence 102, App
40	1130	60.1	354	4	US-10-440-523-102	Sequence 102, App
41	1130	60.1	354	4	US-10-461-925-102	Sequence 102, App
42	1125	59.8	345	4	US-10-146-772-334	Sequence 334, App
43	1125	59.8	345	4	US-10-241-742-334	Sequence 334, App
44	1125	59.8	345	4	US-10-440-523-334	Sequence 334, App
45	1125	59.8	345	4	US-10-440-523-334	Sequence 334, App

ALIGNMENTS

RESULT 1  
US-09-751-299-2  
Sequence 2, Application US/09751299  
Patent No. US20020012974A1  
GENERAL INFORMATION:  
APPLICANT: Madder, Mark  
APPLICANT: Weiner, David P.  
TITLE OF INVENTION: METHODS FOR PRODUCING ENANTIOMERICALLY PURE  
FILE REFERENCE: DIVER1440-2  
CURRENT APPLICATION NUMBER: US/09/751,299  
CURRENT FILING DATE: 2000-12-29  
PRIOR APPLICATION NUMBER: 60/254,414  
PRIOR FILING DATE: 2000-12-07  
PRIOR APPLICATION NUMBER: 60/173,609  
PRIOR FILING DATE: 1999-12-29  
NUMBER OF SEQ ID NOS: 4  
SOFTWARE: Patentin Ver. 2.1  
SEQ ID NO 2  
LENGTH: 346  
TYPE: PRT  
ORGANISM: Unknown Organism  
FEATURE:  
OTHER INFORMATION: Description of Unknown Organism: Obtained from an  
US-09-751-299-2  
OTHER INFORMATION: environmental sample  
Alignment Scores:  
Pred. No.: 6.43e-123  
Score: 1806.00  
Percent Similarity: 100.0%  
Best Local Similarity: 100.0%  
Query Match: 96.0%  
DB: 3  
Length: 346  
Matches: 346  
Conservative: 0  
Mismatch: 0  
Indels: 0  
Gaps: 0

US-09-751-299-1 (1-1041) x US-09-751-299-2 (1-346)

QY 1 ATGTGGAGCCCATGACGAATGATCGCGCGCGGTGACGCGCGCGGTGTCCTC 60  
Db 1 MetSerGIuPromeTThrLyTy-rArgGLyAlaAlaValGlnAlaAlaProValPheLeu 20  
QY 61 GATTTCGACCCGACAGTCGAGAAAGCCGATCGGCTGATCGACGAGCGGCCACAGAC 120  
Db 21 AspLeuAspArgThrValGluValAlaAlaLeuLeuLeuGlnAlaAlaValGlnAsp 40  
QY 121 GTGGCGCTGATCGATTCCGAGACTTGATTCGCGGCTATCCCTTTGGATTGCGTG 180  
Db 41 ValArgLeuLeuAlaPheProGluThrTrpIleProGlyTyPProPheTrpIleTrpLeu 60  
QY 181 GCGCGCGCGCGCTTGGGCGATGCGCTTGTCGACGCGCTATTTCGAAATTCGCTGTCGC 240  
Db 61 G1yAlaProAlaIrrpGlyMetArgPheValGlnArgTyPheGluAsnSerLeuValArg 80  
QY 241 GCGAGCAGCAGTGGCGCGCGCTTGGCGGATGCGCGCGCGCGCGCATGATGTCGTG 300  
Db 81 G1ySerTySglnTrpGlnAlaLeuAlaAspAlaAlaArgArgHISgIyMetHISValVal 100  
QY 301 GCGCGCTATAGCGAGCGCGCGCGCGCGCGCTTATATGCGCGCGCGCATCTTCGCGCCC 360  
Db 101 AlagIyTySerSerGluArgAlaGlyGlySerLeuTyMetGlyGlnAlaAlaIlePheGlyPro 120  
QY 361 GATGGGAGATGATTCGCGCGCGCGCGCGCGCGCTGACAGCTTACCCATGCGCGAGCGCACCGTGC 420  
Db 121 AspP1yAspLeuLeuAlaAlaArgArgLySLeuTySProThHISAlaGluArgThrVal 140  
QY 421 TTCGCGCAGGAGAGACGCGACGCATCTTCGCGTGCACGATACCGCGCATCGCGCGCTCGGC 480  
Db 141 PheGlyGluGlyAspGlySerHISLeuAlaValHISAspHISAlaIleGlyArgLeuGly 160  
QY 481 GCGCTCTGTTGCTGGAGGACATTCAGCCATTTGTCGAATACGCATGTACGCCCGCGAC 540  
Db 161 AlaleuTySCTyPTrpGlnHISIleGlnProLeuSerTyTyAlaMetTyAlaAlaAsp 180  
QY 541 GAACAGGTCCACGCTCGCGCGCGCGCGCGCGCTTACGCTTATGCGCGCGCATGCGCTTATGCG 600  
Db 181 GluGlnValHISValAlaSerTrpProSerPheSerLeuTyArgGlyMetAlaIrryAla 200  
QY 601 CTGCGACCGGAGGTCAATACCGCGCGCGCGCGCGCATCTTACGCGCGCGCTGCTGTCAC 660  
Db 201 LeuGlyProGluValAsnThrAlaAlaSerGlnIleTyAlaValGluGlyGlyCysTy 220  
QY 661 GTGCTGGCGCTGCGCGCGCGCGCGCGCGCTTTCGCGGAGATGATTCAGAGTATTTGGTGATACGCC 720  
Db 221 ValLeuAlaSerTySAlaIrrValIrrSerProGluMetIleuValAlaAspThrPro 240  
QY 721 GACAGAGAGATGTTCTCAAGGCGCGCGCGCGCGCTTTCGCAATGATTTTCGCGCGCGCGCG 780  
Db 241 AspLySgIuMetPheLeuTySAlaGlyGlyGlyPheAlaMetIlePheGlyProAspGly 260  
QY 781 CGGCGCTGGCGCGAGCGCGCGCGCGCGCGCGAGACCGAGAGAGAGATGCTGCTCGCGCATATCGAC 840  
Db 261 ArgAlaLeuAlaGluProLeuProGluThrGlnGlnGlyLeuLeuValAlaAspIleAsp 280  
QY 841 CTGCGCATGATCGCTTGGCGCAAGGCGCGCGCGCGCGCGCGCGCGCATATTCAGCGCGCG 900  
Db 281 LeuGlyMetIleAlaLeuAlaTySAlaAlaAlaAspProAlaGlyHISTySISArgPro 300  
QY 901 GACGTAACCGCGGTGCTGCTGATGCACTCCGCGCGCGCGCGCGCTGCTGCTGATGCGCG 960  
Db 301 AspValThrArgLeuLeuLeuAspArgProAlaGlnArgValAlaThrLeuAspAla 320  
QY 961 GCATTTCGAACCGGAAACGAGAGCAAGGCGCGCGCGCGCGCGCTGCGCGGTGGCGGAGAA 1020  
Db 321 AlaPheGluProGlnAsnGluAspLySgIyAspAlaProAlaLeuArgValAlaAlaGln 340  
QY 1021 AGCGCGCGCGCGCGCGAG 1038  
|||||

Db 341 SerAlaAlaAlaGln 346

RESULT 2  
US-10-146-772-384  
Sequence 384, Application US/10146772  
Publication No. US20030124698A1  
GENERAL INFORMATION:  
APPLICANT: Short, Jay  
APPLICANT: Weiner, David  
APPLICANT: Chaplin, Jennifer  
APPLICANT: Chl, Ellen  
APPLICANT: Milam, Aileen  
APPLICANT: Desantis, Grace  
APPLICANT: Madden, Mark  
APPLICANT: Burk, Mark  
TITLE OF INVENTION: Nitrilases  
FILE REFERENCE: Docket No. US20030124698A1 DIV-01US  
CURRENT APPLICATION NUMBER: US/10/146,772  
CURRENT FILING DATE: 2002-05-15  
PRIOR APPLICATION NUMBER: US 60/309,006  
PRIOR FILING DATE: 2001-07-30  
PRIOR APPLICATION NUMBER: US 60/351,336  
PRIOR FILING DATE: 2002-01-22  
PRIOR APPLICATION NUMBER: US 60/300,189  
PRIOR FILING DATE: 2001-06-21  
PRIOR APPLICATION NUMBER: US 09/751,299  
PRIOR FILING DATE: 2000-12-28  
PRIOR APPLICATION NUMBER: US 60/254,414  
PRIOR FILING DATE: 2000-12-07  
PRIOR APPLICATION NUMBER: US 60/173,609  
PRIOR FILING DATE: 1999-12-29  
NUMBER OF SEQ ID NOS: 386  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 384  
LENGTH: 346  
TYPE: PRT  
ORGANISM: Unknown  
FEATURE:  
OTHER INFORMATION: Obtained from an environmental sample  
US-10-146-772-384

Alignment Scores:  
Pred. No.: 6,43e-123 Length: 346  
Score: 1806.00 Matches: 346  
Percent Similarity: 100.0% Conservative: 0  
Best Local Similarity: 100.0% Mismatches: 0  
Query Match: 96.0% Indels: 0  
Gaps: 0  
DB:

US-09-751-299-1 (1-1041) x US-10-146-772-384 (1-346)

QY 1 ATGTGGAGCCCATGACGAATGATCGCGCGCGGTGACGCGCGCGGTGTCCTC 60  
Db 1 MetSerGIuPromeTThrLyTy-rArgGLyAlaAlaValGlnAlaAlaProValPheLeu 20  
QY 61 GATTTCGACCCGACAGTCGAGAAAGCCGATCGGCTGATCGACGAGCGGCCACAGAC 120  
Db 21 AspLeuAspArgThrValGluValAlaAlaLeuLeuLeuGlnAlaAlaValGlnAsp 40  
QY 121 GTGGCGCTGATCGATTCCGAGACTTGATTCGCGGCTATCCCTTTGGATTGCGTG 180  
Db 41 ValArgLeuLeuAlaPheProGluThrTrpIleProGlyTyPProPheTrpIleTrpLeu 60  
QY 181 GCGCGCGCGCGCTTGGGCGATGCGCTTGTCGACGCGCTATTTCGAAATTCGCTGTCGC 240  
Db 61 G1yAlaProAlaIrrpGlyMetArgPheValGlnArgTyPheGluAsnSerLeuValArg 80  
QY 241 GCGAGCAGCAGTGGCGCGCGCTTGGCGGATGCGCGCGCGCGCGCATGATGTCGTG 300  
Db 81 G1ySerTySglnTrpGlnAlaLeuAlaAspAlaAlaArgArgHISgIyMetHISValVal 100  
QY 301 GCGCGCTATAGCGAGCGCGCGCGCGCGCGCTTATATGCGCGCGCGCATCTTCGCGCCC 360  
|||||



Db 101 AAGlYrSerGuaArGlaGlySerLeuTyMetGlyInAlaIlePheGlyPro 120  
QY 361 GATGGGATTCGATTCGGCGCGCGCGCGCAAGCTCAAGCTTCCATGGGAGCGGACCGTG 420  
Db 121 AaPGLyAspLeuIleAlaAlaArgLysLeuYseProThrHsaIaGlnaGThVal 140  
QY 421 TTCGGGAGGAGAGCGGACCGCATCTCGCGGTGACATACCGCATCGGCGCGCTCGG 480  
Db 141 PheGlyGlnGlyAspGlySerHisLeuAlaValHisAspThraIleGlyArgLeuGly 160  
QY 481 GCGCTCTGTTGCTGGAGACATCCAGCATTTGTGAATACCGCATGTACCGCGGAC 540  
Db 161 AlaleuYseCystrpGlnHisIleGlnProLeuSerLysTyraIleMetTyraIleAlaAsp 180  
QY 541 GAACAGGTCCAGTCGCGGTGTGGCGCGCATTCAGCTTCAAGCTTATCGCGGACATGCG 600  
Db 181 GlnGlnAlaHisValAlaAserTrpProSerPheSerLeuTyraGlyMetAlaTyraIa 200  
QY 601 CTCGACCGGAGGTCAATACCGCGGACGAGTTCACCGGTCGAGGCGGCGCTGCTAC 660  
Db 201 LeuGlyProGlnValAsnThrAlaAlaSerGlnIleTyraIaValGlnGlyCystrYr 220  
QY 661 GTCTGGCGGTCTGTGGCGGACGTTTCGCGGAGATGATCAAGGTATTGTGATACGCC 720  
Db 221 ValLeuAlaSerCystrAlaThrValSerProGlnMetIleLysValLeuValAspThrPro 240  
QY 721 GACAGAGATGTTCTTCAAGCGCGGCGGCGGCTTTCCTCATGATTTTCGGGCGCGGAC 780  
Db 241 AspLysGlnMetPheLeuYsaIaGlyGlyPheIleMetIlePheGlyProAspGly 260  
QY 781 CGGCGCTGGCGGACGCGCTCCGCGGAGACGGAAGAGGACGTCGTCGTCGCGCATATGAC 840  
Db 261 ArgAlaLeuAlaGlnProLeuProGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 280  
QY 841 CTCGGGATGATCGCTTGGCGGACGCGCGCGCGCATTCGCGGCGGCGCATATTCAGCG 900  
Db 281 LeuGlyMetIleAlaLeuAlaLysAlaAlaAlaAspProIleGlyHisTrpSerArgPro 300  
QY 901 GAGCTAACCGGCGCTGCTGCTGATTCGACGTCGCGCGGACGCGCTGCTGATGATGCC 960  
Db 301 AspValThrArgLeuLeuLeuAspArgArgProAlaGlnArgValAlaThrLeuAspAla 320  
QY 961 GCATTGGAACCGGAAAACGAGGACGAGGCGGCGCGCGCGCTGCGGCGGCGGAGAA 1020  
Db 321 AlaPheGlnProGlnAsnGlnAspLysGlyAspAlaProIleLeuArgValAlaAlaGln 340  
QY 1021 AGCGCGCGCGCGCGGACG 1038  
Db 341 SerAlaAlaAlaAlaGln 346

RESULT 3  
US-10-241-742-384  
; Sequence 384, Application US/10241742  
; Publication No. US20040002147A1  
; GENERAL INFORMATION:  
; APPLICANT: Short, Jay  
; APPLICANT: Weiner, David  
; APPLICANT: Chaplin, Jennifer  
; APPLICANT: Chi, Ellen  
; APPLICANT: Milan, Aileen  
; APPLICANT: Desantis, Grace  
; APPLICANT: Madden, Mark  
; APPLICANT: Burk, Mark  
; TITLE OF INVENTION: Nitrites  
; FILE REFERENCE: Docket No. US20040002147A1 DIV-013US  
; CURRENT APPLICATION NUMBER: US/10/241,742  
; CURRENT FILING DATE: 2002-09-09  
; PRIOR APPLICATION NUMBER: US/10/146,772  
; PRIOR FILING DATE: 2002-05-15  
; PRIOR APPLICATION NUMBER: US 60/309,006  
; PRIOR FILING DATE: 2001-07-30  
; PRIOR APPLICATION NUMBER: US 60/351,336  
; PRIOR FILING DATE: 2002-01-22

; PRIOR APPLICATION NUMBER: US 60/300,189  
; PRIOR FILING DATE: 2001-06-21  
; PRIOR APPLICATION NUMBER: US 09/751,299  
; PRIOR FILING DATE: 2000-12-28  
; PRIOR APPLICATION NUMBER: US 60/254,414  
; PRIOR FILING DATE: 2000-12-07  
; PRIOR APPLICATION NUMBER: US 60/173,609  
; PRIOR FILING DATE: 1999-12-29  
; NUMBER OF SEQ ID NOS: 386  
; SOFTWARE: FaSTSeq for Windows Version 4.0  
; SEQ ID NO 384  
; LENGTH: 346  
; TYPE: PRN  
; ORGANISM: Unknown  
; OTHER INFORMATION: Obtained from an environmental sample  
; US-10-241-742-384

Alignment Scores:  
Pred. No.: 6,43e-123 Length: 346  
Score: 1806.00 Matches: 346  
Percent Similarity: 100.0% Conservative: 0  
Best Local Similarity: 100.0% Mismatches: 0  
Query Match: 96.0% Indels: 0  
DB: 4 Gaps: 0

US-09-751-299-1 (1-1041) x US-10-241-742-384 (1-346)

QY 1 ATGTCCGAGCCATGACGAATATCCGCGCGCGGTGACGCGCGCGGTGTTCTC 60  
Db 1 MetSerGlnProMetThrLysTyraGlyAlaAlaValGlnAlaProValPheLeu 20  
QY 61 GATCTGACCGGACGATCGAAGAACGATCGCTGATGACGAGCGCGCAAGCAGAC 120  
Db 21 AspleuAspArgThrValGlnYsaIaIleGlyLeuIleGlnGlnAlaLysGlnAsp 40  
QY 121 GTGCGCTGATGATCCGAGACTTGATTCGCGGCTATTCCTTGGATGATGCTG 180  
Db 41 ValArgLeuIleAlaPheProGlnTrpIleProGlyTyraProPheTrpIleTrpLeu 60  
QY 181 GCGCGCGCGCTTGGGCGATGCGCTTCTGACAGCGCTATTCGAGATTCGCTGCGC 240  
Db 61 GlyAlaProAlaTrpIleMetArgPheValGlnArgTyraPheGlnAsnSerLeuValArg 80  
QY 241 GCGAGCAGCAGTGGCGCGCTGCGGAGATGCGCGCGCGCGCGCATGATGCTG 300  
Db 81 GlySerLysGlnTrpGlnAlaLeuAlaAspAlaAlaArgArgHisGlyMetHisVal 100  
QY 301 GCGCGCTATGAGAGCGCGCGCGCGCGCGCTTATATGAGCGCATCTTCGCGCCC 360  
Db 101 AlaGlyTyraSerGlnArgAlaGlyLysSerLeuTyraMetGlyGlnAlaIlePheGlyPro 120  
QY 361 GATGGGATTCGATTCGGCGCGCGCGCGCAAGCTCAAGCTTCCATGGGAGCGGACCGTG 420  
Db 121 AaPGLyAspLeuIleAlaAlaArgLysLeuYseProThrHsaIaGlnaGThVal 140  
QY 421 TTCGGGAGGAGAGCGGACCGCATCTCGCGGTGACATACCGCATCGGCGCGCTCGG 480  
Db 141 PheGlyGlnGlyAspGlySerHisLeuAlaValHisAspThraIleGlyArgLeuGly 160  
QY 481 GCGCTCTGTTGCTGGAGACATCCAGCATTTGTGAATACCGCATGTACCGCGGAC 540  
Db 161 AlaleuYseCystrpGlnHisIleGlnProLeuSerLysTyraIleMetTyraIleAlaAsp 180  
QY 541 GAACAGGTCCAGTCGCGGTGTGGCGCGCATTCAGCTTCAAGCTTATCGCGGACATGCG 600  
Db 181 GlnGlnAlaHisValAlaAserTrpProSerPheSerLeuTyraIleMetTyraIleAlaAsp 200  
QY 601 CTCGACCGGAGGTCAATACCGCGGACGAGTTCACCGGTCGAGGCGGCGCTGCTAC 660  
Db 201 LeuGlyProGlnValAsnThrAlaAlaSerGlnIleTyraIaValGlnGlyCystrYr 220  
QY 661 GTCTGGCGGTCTGTGGCGGACGTTTCGCGGAGATGATCAAGGTATTGTGATACGCC 720

```
Db 221 ValLeuAlaSerCySaIaThrValSerProGluMetIleValLeuValaIaPThrPro 240
QY 721 GACAAGAGATGTTCTTCAAGCGCGCGGCTTTTCCATGATTTTCGGCCGACGCG 780
Db 241 AspLysGluMetPheLeuLysAlaGlyGlyPheAlaMetIlePheGlyProAspGly 260
QY 781 CGGCGCTGGCCGACCGCGCTCCCGAGACCGAAGAGGACTGCTGGTCCGATATGAC 840
Db 261 ArgAlaLeuAlaGluProLeuProGluThrGluGluGlyLeuLeuValaIaAspIleAsp 280
QY 841 CTGGGATGATCGCGTTGGCCAGGCGCGCGCGATCCGGCGGCGCCACTATTACGCGCC 900
Db 281 LeuGlyMetIleAlaLeuAlaLysAlaAlaAspProAlaGlyIleStyIleSerArgPro 300
QY 901 GACGTAACGGCGGTGCTGTGATGACAGCTCCGCGCCAAAGCGCTGTCACGCTTAGTGC 960
Db 301 AspValThrArgLeuLeuLeuAspArgArgProAlaGlnArgValaIleThrLeuAspAla 320
QY 961 GCATTGGAACCGGAACGAGGACCAAGGCGGACCGCGCGCGCTGGCGGTGGCGGAA 1020
Db 321 AlaPheGluProGluAsnGluAspLysGlyAspAlaProAlaLeuArgValaIaIaGln 340
QY 1021 AGCGCCGCGCGCCGCGCAG 1038
Db 341 SerAlaAlaAlaIaGln 346

RESULT 4
US-10-440-523-384
/ Sequence 384, Application US/10440523
/ Publication No. US20040014195A1
/ GENERAL INFORMATION:
/ APPLICANT: Weiner, David
/ APPLICANT: Short, Jay
/ APPLICANT: Chaplin, Jennifer
/ APPLICANT: Chi, Ellen
/ APPLICANT: Milan, Aileen
/ APPLICANT: Desantis, Grace
/ APPLICANT: Madden, Mark
/ APPLICANT: Burk, Mark
/ TITLE OF INVENTION: Nitrlases
/ FILE REFERENCE: Docket No. US20040014195A1 DIV-013US
/ CURRENT APPLICATION NUMBER: US/10/440,523
/ CURRENT FILING DATE: 2003-05-15
/ PRIOR APPLICATION NUMBER: US/10/146,772
/ PRIOR FILING DATE: 2002-05-15
/ PRIOR APPLICATION NUMBER: US 60/309,006
/ PRIOR FILING DATE: 2001-07-30
/ PRIOR APPLICATION NUMBER: US 60/351,336
/ PRIOR FILING DATE: 2002-01-22
/ PRIOR APPLICATION NUMBER: US 60/300,189
/ PRIOR FILING DATE: 2001-06-21
/ PRIOR APPLICATION NUMBER: US 09/751,299
/ PRIOR FILING DATE: 2000-12-28
/ PRIOR APPLICATION NUMBER: US 60/254,414
/ PRIOR FILING DATE: 2000-12-07
/ PRIOR APPLICATION NUMBER: US 60/173,609
/ PRIOR FILING DATE: 1999-12-29
/ NUMBER OF SEQ ID NOS: 386
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 384
/ LENGTH: 346
/ TYPE: PRT
/ ORGANISM: Unknown
/ FEATURE:
/ OTHER INFORMATION: Obtained from an environmental sample
US-10-440-523-384

Alignment Scores:
Pred. No.: 6,43e-123 Length: 346
Score: 1806.00 Matches: 346
Percent Similarity: 100.0% Conservative: 0
Best Local Similarity: 100.0% Mismatches: 0
```

```
Query Match: 96.0% Indels: 0
DB: 4 Gaps: 0
US-09-751-299-1 (1-1041) x US-10-440-523-384 (1-346)
QY 1 ATGTGGAGGCCCATATGACGAAGTATGCGCGCGCGGCGGACAGGCCCGGCTGTTCCCTC 60
Db 1 MetSerGluProMetThrLysTyrArgGlyAlaAlaValaGlnAlaIleProValaPheLeu 20
QY 61 GATTCGACCCGACAGTGCAGAAAGCGATCGGCTGATCGACAGCGGCCCAAGAGAC 120
Db 21 AspLeuAspArgThrValGlyLysAlaIleGlyLeuIleGluGlnAlaAlaLysGlnAsp 40
QY 121 GTGGCGCTGTATGCAATCCCAAGACTTGATCCCGGCTATCCCTTTGGATATAGCGCTG 180
Db 41 ValArgLeuIleAlaPheProGluThrTrpIleProGluTyrProPheTrpIleThrLeu 60
QY 181 GCGCGCGCGGCTTGGGGCGATGCGCTTCGTCAGCGCTATTGCGAATTCGCTGCGCGC 240
Db 61 GlyAlaProAlaIleTrpGlyMetArgPheValGlnArgTyrPheGluAsnSerLeuValArg 80
QY 241 GCGACGACAGCACTGGCGAGCGCTGGCGGATGCGCGCGCGCCGCGACATGCAATGCTG 300
Db 81 GlySerLysGlnTrpGlnAlaLeuAlaAspAlaAlaArgAlaAlaArgAlaGlyMetIleVal 100
QY 301 GCGGCTATPAGGAGCGCGCGCGCGCGAGCGCTTATPAGGCGAGCGATCTTCGCGCCC 360
Db 101 AlaGlyTyrSerGluArgAlaGlyLysSerLeuTyrMetGlyGlnAlaIlePheGlyPro 120
QY 361 GATGCGCATCTGATGCGCGCGCGCGCGCAAGCTCAAGCTTACCATGCGAGCGCAACGCTG 420
Db 121 AspGlyAspLeuIleAlaAlaArgAlaGlyLysLeuLysProThrIleAlaGlyThrVal 140
QY 421 TTCGGGAGGAGGAGAGCGGACGACATCTCGCGGAGCGAATACCGCATGCGGCGCTCGGC 480
Db 141 PheGlyGluGlyAspGlySerIleValaIleAspThrAlaIleGlyArgLeuGly 160
QY 481 GCGCTCTGTGTGGGAGACACATCCAGCATTTGGAATACGCCATGTACGCGCCGAC 540
Db 161 AlaLeuCyGcySerTrpGlnIleIleGlnProLeuSerLysTyrAlaMetTyrAlaIleAsp 180
QY 541 GACAAGGTCCACGTGCGCGCTGTGGCGGACGTTCAAGCTTATCGGCGCATGCGCTATGCG 600
Db 181 GlnGlnValaIleValaIleSerTrpProSerPheSerLeuTyrArgGlyMetAlaTyrAla 200
QY 601 CTGGACCGGAGATCAATACCGCGCGCAAGCGCATGCTAGCGGCTGAGAGCGGCTGTAC 660
Db 201 LeuGlyProGluValaIleThrAlaIleSerGlnIleTyrAlaValaGluGlyCySty 220
QY 661 GTGCTGGCGTCTGGCGGACCGCTTTCGCGGATGATCAAGTATGATGATGATGATGATGATG 720
Db 221 ValLeuAlaSerCySaIaThrValSerProGluMetIleLysValLeuValaIaPThrPro 240
QY 721 GACAAGAGATGTTCTTCAAGCGCGCGGCTTTTCCATGATTTTCGGCCGACGCG 780
Db 241 AspLysGluMetPheLeuLysAlaGlyGlyPheAlaMetIlePheGlyProAspGly 260
QY 781 CGGCGCTGGCCGACCGCGCTCCCGAGACCGAAGAGGACTGCTGGTCCGATATGAC 840
Db 261 ArgAlaLeuAlaGluProLeuProGluThrGluGluGlyLeuLeuValaIaAspIleAsp 280
QY 841 CTGGGATGATCGCGTTGGCCAGGCGCGCGCGATCCGGCGGCGCCACTATTACGCGCC 900
Db 281 LeuGlyMetIleAlaLeuAlaLysAlaAlaAspProAlaGlyIleStyIleSerArgPro 300
QY 901 GACGTAACGGCGGTGCTGTGATGACAGCTCCGCGCCAAAGCGCTGTCACGCTTAGTGC 960
Db 301 AspValThrArgLeuLeuLeuAspArgArgProAlaGlnArgValaIleThrLeuAspAla 320
QY 961 GCATTGGAACCGGAACGAGGACCAAGGCGGACCGCGCGCGCTGGCGGTGGCGGAA 1020
Db 321 AlaPheGluProGluAsnGluAspLysGlyAspAlaProAlaLeuArgValaIaIaGln 340
```

QY 1021 AGCGCCGCGCGCGCAG 1038  
Db 341 SerAlaAlaAlaAlaGln 346

RESULT 5  
US-10-440-503-384  
; Sequence 384, Application US/10440503  
; Publication No. US20040038419A1  
; GENERAL INFORMATION:  
; APPLICANT: Weiner, David Paul  
; APPLICANT: Chaplin, Jennifer Ann  
; APPLICANT: Chi, Ellen  
; APPLICANT: Milan, Aileen  
; APPLICANT: Desantis, Grace  
; APPLICANT: Burk, Mark U.  
; APPLICANT: McQuaid, Jeffrey  
; APPLICANT: Stege, Justin  
; TITLE OF INVENTION: ASSAYS AND KITS FOR DETECTING THE PRESENCE OF  
; TITLE OF INVENTION: NITRILES AND/OR CYANIDE  
; FILE REFERENCE: 09010-900001  
; CURRENT APPLICATION NUMBER: US/10/440, 503  
; CURRENT FILING DATE: 2003-05-15  
; PRIOR APPLICATION NUMBER: US 60/380, 737  
; PRIOR FILING DATE: 2002-05-15  
; NUMBER OF SEQ ID NOS: 386  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 384  
; LENGTH: 346  
; TYPE: PRT  
; ORGANISM: Unknown  
; OTHER INFORMATION: Obtained from an environmental sample  
US-10-440-503-384

Alignment Scores:  
Pred. No.: 6,43e-123 Length: 346  
Score: 1806.00 Matches: 346  
Percent Similarity: 100.0% Conservative: 0  
Best Local Similarity: 100.0% Mismatches: 0  
Query Match: 96.0% Indels: 0  
DB: 4 Gaps: 0

US-09-751-299-1 (1-1041) x US-10-440-503-384 (1-346)

QY 1 ATGTGAGAGCCCATGACGAAGTATCGCGCGCGGTGACGCGCGCGGTCTCTC 60  
Db 1 MetSerGluProMetThrLysrYrArgGlyAlaAlaValGlnAlaAlaProValPheLeu 20

QY 61 GATGTGACCGGCAGATCGAGAAAGCATCGGCTGATCGAGCAGCGCGCCAGCAGAC 120  
Db 21 AspLeuAspArgThrValGlnLysAlaAlaLeuLeuLeuGlnAlaAlaLysGlnAsp 40

QY 121 GTGGCGCTGATCGGATTCGCCAGACTTGATCCCGGCTATCCCTTTGGATGGGTG 180  
Db 41 ValArgLeuIleAlaPheProGluThrTrpIleProGluYrProPheTrpIleTrpLeu 60

QY 181 GCGCGCGCGCTTGGGCGATCGCTTGTCTCCAGCGCTATTGCGAATTCGCTGTGCGC 240  
Db 61 GlyAlaProAlaTrpGlyMetArgPheValGlnArgYrPheGlnuAnSerLeuValArg 80

QY 241 GCGAGCAGACGATGGGAGCGCTTGGCGGATCGGCGCGCGCGCGCGCATGCTGTG 300  
Db 81 GlySerLysGlnTrpGlnAlaLeuAlaAspAlaAlaArgAlaGlnLysMetHsValVal 100

QY 301 GCCGCTATAGCGAGCGCGCGCGCGAGCTTATATGAGCGAGCATTCGAGCGCC 360  
Db 101 AlaGlyYrSerGlnuAlaGlnAlaGlyLysLeuYrMetGlyGlnAlaAlaPheGlyPro 120

QY 361 GATGCGCATGTGATCGCGCGCGCGCGAGCTCAAGCTCAAGCTCAAGCTCGAGCGACCGTG 420  
Db 121 AspGlyAspLeuIleAlaAlaArgArgLysLeuYrProThrHsAlaGlnArgThrVal 140

QY 421 TTCGCGAGGAGACGCGCAGCATCTCGCGGTGACATACCGGCATCGGCGCGCTCGGC 480

Db 141 PheGlyGlnLysAspGlySerHsLeuAlaValHsAspThrAlaAlaLeuArgLeuGly 160

QY 481 GCGCTGTGTTCTGGAGACATTCACGCGCATTCGAAATACGCGCATGACGCGCGAC 540  
Db 161 AlaLeuCysCysrTrpGlnHsIleGlnProLeuSerYrThrAlaMetCyrAlaAlaAsp 180

QY 541 GAACAGTCCACGTCGCGGTGTGGCGCGCATTCAGCTCTATCGCGCGCATGCGCTATGCG 600  
Db 181 GlnGlnValHsValAlaSerTrpProSerPheSerLeuYrArgGlyMetAlaYrAla 200

QY 601 CTCGAGCAGAGGTCAATACCGCGCAGCGCAGATCAAGCTGACGCGCGCGCTGTAC 660  
Db 201 LeuGlyProGlnValAlaThrAlaAlaSerGlnIleYrAlaValGlnGlyCysYr 220

QY 661 GTGCTGGCGTCTGCGCGACCGTTTCGCGAGATGATCAAGTATTGTGATACGCC 720  
Db 221 ValLeuAlaSerCysAlaThrValSerProGluMetCileYsValLeuValAspThrPro 240

QY 721 GACAGAGATGTTCTCAAGCGCGCGCGGCTTTCCATGATTTTCGCGCGCGACGCC 780  
Db 241 AspLysGlnMetPheLeuYsAlaGlyGlyPheAlaMetIlePheGlyProAspGly 260

QY 781 CGCGCGCTGGCGCGCGCTCCGAGACCGAAGAGGAGTGTGTGCGCGCATATGAC 840  
Db 261 ArgAlaLeuAlaGlnuProLeuProGluThrGlnGlnuLeuLeuValAlaAspLeaAsp 280

QY 841 CTCGCGATGATCGGTGGCGCAAGCGCGCGCGCATTCGCGCGCGCATTCACGCGCC 900  
Db 281 LeuGlyMetIleAlaLeuAlaLysAlaAlaAlaAspProAlaGlnHsYrSerArgPro 300

QY 901 GACGTAAAGCGGCTGCTGTGATCGAGTCGCGCGCGCAAGCGCTGTGATGATGCC 960  
Db 301 AspValThrArgLeuLeuLeuAspArgArgProAlaGlnArgValAlaThrLeuAspAla 320

QY 961 GCATTGGAACCGGAAACGAGACGAGCGCGCGCGCGCGCGGTGGTGGCGGAA 1020  
Db 321 AlaPheGlnuProGlnuGlnuAspLysGlyAspAlaProAlaLeuArgValAlaGln 340

QY 1021 AGCGCCGCGCGCGCAG 1038  
Db 341 SerAlaAlaAlaAlaGln 346

RESULT 6  
US-10-461-925-384  
; Sequence 384, Application US/10461925  
; Publication No. US20040053378A1  
; GENERAL INFORMATION:  
; APPLICANT: Mark J. Burk  
; APPLICANT: Desantis, Grace  
; APPLICANT: Morgan, Brian  
; APPLICANT: Zhu, Zoulin  
; TITLE OF INVENTION: PROCESSES FOR MAKING (R)-ETHYL 4-CYANO-3-HYDROXYBUTYRIC ACID  
; FILE REFERENCE: 09010-270001  
; CURRENT APPLICATION NUMBER: US/10/461, 925  
; CURRENT FILING DATE: 2003-06-13  
; PRIOR APPLICATION NUMBER: US 60/389, 317  
; PRIOR FILING DATE: 2002-06-13  
; PRIOR APPLICATION NUMBER: US 60/392, 944  
; PRIOR FILING DATE: 2002-06-28  
; NUMBER OF SEQ ID NOS: 386  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 384  
; LENGTH: 346  
; TYPE: PRT  
; ORGANISM: Unknown  
; FEATURE:  
; OTHER INFORMATION: Obtained from an environmental sample  
US-10-461-925-384

Alignment Scores:  
Pred. No.: 6,43e-123 Length: 346  
Score: 1806.00 Matches: 346

Percent Similarity: 100.0%  
Best Local Similarity: 100.0%  
Query Match: 96.0%  
DB: 4  
Conservative: 0  
Mismatches: 0  
Indels: 0  
Gaps: 0

US-09-751-299-1 (1-1041) x US-10-461-925-384 (1-346)

```
QY 1 ATGTGGAGCCCATGACGAAGTATCGCGCGCGGCGGTGACAGCGCGCGGTGTTCTC 60
DB 1 MetSerLupProMetThrLysTyrArgGlyAlaIalValGlnAlaIalProValPheLeu 20
QY 61 GATCTGACCGGACAGCGGAAAGGATCGGCTGATTCAGAGCGGCGGCGGCGGCGGAG 120
DB 21 AspLeuAspArgThrValGlnLysAlaIleGlyLeuIleGlnGlnAlaIalysGlnAsp 40
QY 121 GTGGCGCTGATCGCATTCAGAGACTTGATTCGCGCTATCCCTTTGGATGAGCTG 180
DB 41 ValArgLeuIleAlaPheProGlnThrTyrIleProGlyTyrProPheTyrIleTyrPhe 60
QY 181 GGGCGCGCGCTTGGGCGCATGCGCTTGTGTCAGCGCTATTCAGAAATTCGCTGCGC 240
DB 61 GlyAlaProAlaTyrGlyMetArgPheValGlnArgTyrPheGlnAsnSerLeuValArg 80
QY 241 GGCAGCAGAGATGCGAGCGCTGCGAGATGCGGCGCGCGCGCGCATGATGTCGTG 300
DB 81 GlySerLysGlnTyrGlnAlaLeuAlaAspAlaAlaArgArgThrGlyMetHisValVal 100
QY 301 GCGGCGCTATGCGAGCGCGCGCGCGCGCGCTTATATGCGGCGAGCGATCTTGGCGCC 360
DB 101 AlAGlyTyrSerGlnArgAlaGlyGlySerLeuTyrMetGlyGlnAlaIlePheGlyPro 120
QY 361 GATGCGCATGATCGCGCGCGCGCGCGAGCTCAAGCTCAATGCGGAGCGGCGCGCGT 420
DB 121 AspGlyAspLeuIleAlaAlaArgGlyLeuIleProThrHisAlaGlnAlaGlnVal 140
QY 421 TTGCGGAGGAGAGACGCGACCGCATCTTCGCGTGCAGATACCGCATGCGGCGCTGCGC 480
DB 141 PheGlyGlnIleAspGlySerHisLeuAlaValHisAspThrAlaIleGlyArgLeuGly 160
QY 481 GCGCTGTTGCTGGGAGCATTCGAGCCATTCGTAATACCAATGTAACGCGCGCGAG 540
DB 161 AlLeuGlyCysTyrGlnHisIleGlnProLeuSerLysTyrAlaMetCylAlaAlaAsp 180
QY 541 GAAAGGTCACGTCGCGGTGCGGCGAGCTTCAGCCTCTATCGCGCATGAGCTTATGCG 600
DB 181 GlnGlnValHisValAlaSerTyrProSerPheSerLeuTyrArgGlyMetAlaTyrAla 200
QY 601 CTGCGACCGGAGGTCATTAACGCGCGAGCAAGCAATCTACCGCGGTGCGGCGCTGCTAC 660
DB 201 LeuGlyProGlnValAsnThrAlaAlaSerGlnIleTyrAlaValGlnGlyCysTyr 220
QY 661 GTGCTGCGCTGTCGCGAGCCGTTTCGCGAGATGATCAAGGATATTGATGATACGCC 720
DB 221 ValLeuAlaSerCysAlaThrValSerProGlnMetIleLysValLeuValAspThrPro 240
QY 721 GACAAAGAGATGTTCTCAAGCGCGCGCGGCTTTTGCATGATTTTGGGCGCGAGCGC 780
DB 241 AspLysGlnMetPheLeuLysAlaGlyGlyPheAlaMetIlePheGlyProAspGly 260
QY 781 CGGCGCTGCGCGAGCGCTCCCGGAGACCGAGAGAGGAGCTGCTGGCGCGCATTCAGC 840
DB 261 ArgAlaLeuAlaGlnLupProLeuProGlnThrGlnGlnLysLeuValAlaAspIleAsp 280
QY 841 CTGCGCATGATGCGCTTGGCGCAAGCGCGCGCGCGATTCGCGGCGCGCATATTGACGCC 900
DB 281 LeuGlyMetIleAlaLeuAlaLysAlaAlaAspProAlaGlyHisTyrSerArgPro 300
QY 901 GACGTAACGCGGCTGCTGATGATGAGTCGCGCGCGCGCGCATGCGCTGATGATGCGC 960
DB 301 AspValThrArgLeuLeuLeuAspArgProAlaGlnArgValAlaThrLeuAspAla 320
QY 961 GCATTGAAACCGAAACGAGGACGAGGCGCGCGCGCTGCGCTGCTGCGGCGGAA 1020
```

DB 321 AlaPheGlnProGlnAsnGlnAspLysGlyAspAlaProAlaLeuArgValAlaGln 340  
QY 1021 AGCGCGCGCGCGCGCGCGAG 1038  
DB 341 SerAlaAlaAlaGln 346

## RESULT 7

```
US-10-146-772-34
; Sequence 34, Application US/10146772
; Publication No. US20030124698A1
; GENERAL INFORMATION:
; APPLICANT: Short, Day
; APPLICANT: Weiner, David
; APPLICANT: Chaplin, Jennifer
; APPLICANT: Chi, Ellen
; APPLICANT: Milan, Allen
; APPLICANT: Desantis, Grace
; APPLICANT: Madden, Mark
; APPLICANT: Burk, Mark
; TITLE OF INVENTION: Nitrites
; FILE REFERENCE: Docket No. US20030124698A1 DIV-013US
; CURRENT APPLICATION NUMBER: US/10/146,772
; PRIOR FILING DATE: 2002-05-15
; PRIOR APPLICATION NUMBER: US 60/309,006
; PRIOR FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US 60/351,336
; PRIOR FILING DATE: 2002-01-22
; PRIOR APPLICATION NUMBER: US 60/300,189
; PRIOR FILING DATE: 2001-06-21
; PRIOR APPLICATION NUMBER: US 09/751,299
; PRIOR FILING DATE: 2000-12-28
; PRIOR APPLICATION NUMBER: US 60/254,414
; PRIOR FILING DATE: 2000-12-07
; PRIOR APPLICATION NUMBER: US 60/173,609
; PRIOR FILING DATE: 1999-12-29
; NUMBER OF SEQ ID NOS: 386
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 34
; LENGTH: 341
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Obtained from an environmental sample
US-10-146-772-34
```

## Alignment Scores:

Pred. No.:	5,11e-98	Length:	341
Score:	1464.00	Matches:	272
Percent Similarity:	90.1%	Conservative:	28
Best Local Similarity:	81.7%	Mismatch:	33
Query Match:	77.8%	Indels:	0
DB:	4	Gaps:	0

US-09-751-299-1 (1-1041) x US-10-146-772-34 (1-341)

```
QY 1 ATGTGGAGCCCATGACGAAGTATCGCGCGCGGCGGTGACAGCGCGCGGTGTTCTC 60
DB 1 MetLeuSerProValThrGlnTyrArgAlaAlaValGlnAlaAlaProSerPheLeu 20
QY 61 GATCTGACCGGACAGCGTGGAGAAAGGATCGGCTGATTCAGAGCGGCGGCGGCGGAG 120
DB 21 AspLeuAspArgThrValGlnLysThrIleAlaIleGlnGlnAlaIalGlnAsp 40
QY 121 GTGGCGCTGATGCAATTCAGAGACTTGATTCGCGCTATCCCTTTGGATGAGCTG 180
DB 41 ValArgLeuIleAlaPheProGlnThrTyrIleProGlyTyrProPheTyrIleTyrPhe 60
QY 181 GCGGCGCGCGCTTGGGCGCATGCGCTTGTGTCAGCGCTATTCAGAAATTCGCTGCGC 240
DB 61 GlySerProAlaTyrGlnMetArgPheValGlnArgTyrPheGlnAsnSerLeuValArg 80
QY 241 GGCAGCAGCAGTGCAGCGCTGCGAGATGCGGCGCGCGCGCGCATGATGTCGTG 300
```

```
Db      81 GlySerIsgInTrpAenAlaIleAlaAspAlaIaArgHisArgMetThrValVal 100
Qy      301 GCCGGCTATAGCGCGCGCGCGCGCGCGCTCTATATGGCCCGGAGATCTTGGGCCC 360
Db      101 ValGlyPheSerGluValArgIaGlyGlySerLeuTyMetGlyIaAlaIlePheGlyPro 120
Qy      361 GATGGCATGTGATCGCGCGCGCGCGCGCGCATGCAAGCTCAAGCTCAAGCTCAAGCTG 420
Db      121 GluGlyGluLeuIleAlaIaArgArgIleuLeuSerProThrIleAlaGluArgThrVal 140
Qy      421 TTCGGCAGGAGAGACGGCAGCATCTCGCGGTGACGATACCGCATCGGCGCGCTCGGC 480
Db      141 PheGlyGluGlyAspGlySerHisIleuAlaValIleGluThrGlyValGlyArgIleGly 160
Qy      481 GCGCTCTGTGTCGGGAGACATCCAGCATTCGCGAAATACCGCATGATGACCGCGGAC 540
Db      161 AlaLeuCySerTrpGluHisIleGluProLeuSerIleTyAlaMetTyAlaAlaAsn 180
Qy      541 GAACAGGTCCACGTCGCGCGCGCGCGCGCGCTTACGCTTACGCGCGCATGCGCTATGCG 600
Db      181 GluGluValHisValAlaIaSerTrpProCyPheSerLeuTyArgIleMetAlaTyAla 200
Qy      601 CTCGACCGGAGGTCAATACCGCGCGCAAGCCAGATCTACGCGGTGAGGCGGCTGCTAC 660
Db      201 LeuGlyProGluValAsnThrAlaIaIaSerGluValTyAlaValGluGlyGlyCysTyr 220
Qy      661 GTGCTGCGCTGTCGCGCGCGCGCGCGCGCGCGCGCATGATCAAGTATGTGTGATACGCC 720
Db      221 ValLeuAlaSerCySerLeuValIaIleThrProGluIleLeuValIleAspThrPro 240
Qy      721 GACAGAGAGATGTTCTCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 780
Db      241 AspIleuLeuProLeuLeuLeuAlaGlyGlyIlePheSerMetIlePheGlyProAspGly 260
Qy      781 CGCGCGCGCGCGCGCGCGCGCTCCGCGAGACCGAGAGGAGACTGTGTGCGCGATATGAC 840
Db      261 ArgAlaLeuAlaGluProLeuProGluThrGluGluGlyLeuValThrAlaGluIleAsp 280
Qy      841 CTGCGCATGATCGCGCTTGGCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 900
Db      281 LeuGlyAlaIleAlaLeuAlaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIaIa 300
Qy      901 GAGGTACGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 960
Db      301 AspValIleThrArgLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 320
Qy      961 GCATTGCAACCGCAAAACGAGACGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 999
Db      321 ArgPheGluValValGluIleSerGluGluAlaGluProPro 333

RESULT 8
US-10-241-742-34
; Sequence 34, Application US/10241742
; Publication No. US20040002147A1
; GENERAL INFORMATION:
; APPLICANT: Short, Jay
; APPLICANT: Weiner, David
; APPLICANT: Chaplin, Jennifer
; APPLICANT: Chi, Ellen
; APPLICANT: Milan, Aileen
; APPLICANT: Desantis, Grace
; APPLICANT: Madden, Mark
; APPLICANT: Burk, Mark
; TITLE OF INVENTION: NitriIaees
; FILE REFERENCE: Docket No. US20040002147A1 DIV-013US
; CURRENT APPLICATION NUMBER: US/10/241,742
; CURRENT FILING DATE: 2002-09-09
; PRIOR APPLICATION NUMBER: US/10/146,772
; PRIOR FILING DATE: 2002-05-15
; PRIOR APPLICATION NUMBER: US 60/309,006
; PRIOR FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US 60/351,336
; PRIOR FILING DATE: 2002-01-22
```

```
; PRIOR APPLICATION NUMBER: US 60/300,189
; PRIOR FILING DATE: 2001-06-21
; PRIOR APPLICATION NUMBER: US 09/751,299
; PRIOR FILING DATE: 2000-12-28
; PRIOR APPLICATION NUMBER: US 60/254,414
; PRIOR FILING DATE: 2000-12-07
; PRIOR APPLICATION NUMBER: US 60/173,609
; PRIOR FILING DATE: 1999-12-29
; NUMBER OF SEQ ID NOS: 386
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 34
; LENGTH: 341
; TYPE: PRF
; ORGANISM: Unknown
; OTHER INFORMATION: Obtained from an environmental sample
US-10-241-742-34

Alignment Scores:
Pred. No.: 5,11e-98 Length: 341
Score: 1464.00 Matches: 272
Percent Similarity: 90.1% Conservative: 28
Best Local Similarity: 81.7% Mismatches: 33
Query Match: 77.8% Indels: 0
DB: 4 Gaps: 0

US-09-751-299-1 (1-1041) x US-10-241-742-34 (1-341)
Qy      1 ATGTGAGAGCCATGACGAATATCGCGCGCGCGGTGAGGCGCGCGGTGTTCTC 60
Db      1 MetLeuSerProValThrGlnTyArgAlaIaIaValGluIaIaIaProSerPheLeu 20
Qy      61 GATCTGACCGCGCACTGTCGAGAAAGCATGCGCTGATCGACGAGCGCGCGCAAGCAGAC 120
Db      21 AspLeuAspArgThrValGluTyThrIleAlaIleIleGluGluIaIaIaGluGluAsp 40
Qy      121 GTGCGCTGATGTCATCCGAGACTGGATTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 180
Db      41 ValArgLeuIleAlaPheProGluIleThrIleProGlyTyProLeuIleThrIlePro 60
Qy      181 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 240
Db      61 GlySerProAlaIleTrpGlyMetAlaGlyPheValGlnArgTyPheGluIleSerLeuValArg 80
Qy      241 GGCAGACGACGTGCGAGCGCTTGGCGGATCGCGCGCGCGCGCGCGCGCGCGCGCGCG 300
Db      81 GlySerIsgInTrpAenAlaIleAlaAspAlaIaArgHisArgMetThrValVal 100
Qy      301 GCCGGCTATAGCGCGCGCGCGCGCGCGCTCTATATGGCCCGGAGATCTTGGGCCC 360
Db      101 ValGlyPheSerGluValArgIaGlyGlySerLeuTyMetGlyIaAlaIlePheGlyPro 120
Qy      361 GATGGCATGTGATCGCGCGCGCGCGCGCGCATGCAAGCTCAAGCTCAAGCTCAAGCTG 420
Db      121 GluGlyGluLeuIleAlaIaArgArgIleuLeuSerProThrIleAlaGluArgThrVal 140
Qy      421 TTCGGCAGGAGAGACGGCAGCATCTCGCGGTGACGATACCGCATCGGCGCGCTCGGC 480
Db      141 PheGlyGluGlyAspGlySerHisIleuAlaValIleGluThrGlyValGlyArgIleGly 160
Qy      481 GCGCTCTGTGTCGGGAGACATCCAGCATTCGCGAAATACCGCATGATGACCGCGGAC 540
Db      161 AlaLeuCySerTrpGluHisIleGluProLeuSerIleTyAlaMetTyAlaAlaAsn 180
Qy      541 GAACAGGTCCACGTCGCGCGCGCGCGCGCGCTTACGCTTACGCGCGCATGCGCTATGCG 600
Db      181 GluGluValHisValAlaIaSerTrpProCyPheSerLeuTyArgIleMetAlaTyAla 200
Qy      601 CTCGACCGGAGGTCAATACCGCGCGCAAGCCAGATCTACGCGGTGAGGCGGCTGCTAC 660
Db      201 LeuGlyProGluValAsnThrAlaIaIaSerGluValTyAlaValGluGlyGlyCysTyr 220
Qy      661 GTGCTGCGCTGTCGCGCGCGCGCGCGCGCGCGCGCGCATGATCAAGTATGTGTGATACGCC 720
```

```
Db 221 ValLeuAlaSerCysLeuValValThrProGluIleLeuValLeuIleAspThrPro 240
Qy 721 GACAAAGAGATGTTCTCTCAAGCGCGCGCGGCTTTCCATGATTTTCGGGCCGACGGC 780
Db 241 AspLysGluProLeuLeuLeuAlaGlyGlyPheSerMetIlePheGlyProAspGly 260
Qy 781 CGGCGCTGGCGGCGCTCCCGAGACCGAAGAGGAGCTGGCTGGCGCGCATATGAC 840
Db 261 ArgAlaLeuAlaGlnProLeuProGluThrGluGluLeuValThrAlaGluIleAsp 280
Qy 841 CTCGGCATGATCGCGTTGGCCCAAGCGCGCGCGATCCGCGCGCATATTCACGCGCC 900
Db 281 LeuGlyAlaIleAlaLeuAlaValAlaAlaAlaAspProAlaGlyHisTyrAlaArgPro 300
Qy 901 GAGTAAACGGCGCTGGCTGGATTCGACGTCGCGCCCAAGCGCTGGCATGATGCC 960
Db 301 AspValThrArgLeuLeuLeuAsnProArgProAlaAlaArgValGluAlaGluGlyPro 320
Qy 961 GCATTGCAACCGCAAAACGAGACAAAGGCGACGCGCC 999
Db 321 ArgPheGluValValGlnSerGluGlnAlaGluProPro 333
```

## RESULT 9

```
US-10-440-523-34
; Sequence 34, Application US/10440523
; Publication No. US20040014195A1
; GENERAL INFORMATION:
; APPLICANT: Short, Jay
; APPLICANT: Weiner, David
; APPLICANT: Chai, Ellen
; APPLICANT: Milan, Allen
; APPLICANT: Desantis, Grace
; APPLICANT: Madden, Mark
; APPLICANT: Burk, Mark
; TITLE OF INVENTION: Nitrlases
; FILE REFERENCE: Docket No. US20040014195A1 DIV-013US
; CURRENT APPLICATION NUMBER: US/10/440,523
; PRIOR FILING DATE: 2003-05-15
; PRIOR APPLICATION NUMBER: US/10/146,772
; PRIOR FILING DATE: 2002-05-15
; PRIOR APPLICATION NUMBER: US 60/309,006
; PRIOR FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US 60/351,336
; PRIOR FILING DATE: 2002-01-22
; PRIOR APPLICATION NUMBER: US 60/300,189
; PRIOR FILING DATE: 2001-06-21
; PRIOR APPLICATION NUMBER: US 09/751,299
; PRIOR FILING DATE: 2000-12-28
; PRIOR APPLICATION NUMBER: US 60/254,414
; PRIOR FILING DATE: 2000-12-07
; PRIOR APPLICATION NUMBER: US 60/173,609
; PRIOR FILING DATE: 1999-12-29
; NUMBER OF SEQ ID NOS: 386
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 34
; LENGTH: 341
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Obtained from an environmental sample
US-10-440-523-34
```

## Alignment Scores:

```
Pred. No.: 5.11e-98 Length: 341
Score: 1464.00 Matches: 272
Percent Similarity: 90.1% Conservative: 28
Best Local Similarity: 81.7% Mismatches: 33
Query Match: 77.8% Indels: 0
DB: 4 Gaps: 0
```

US-09-751-299-1 (1-1041) x US-10-440-523-34 (1-341)

```
Qy 1 ATGTCGAGCCCATGACGAAGTATCGCGCGCGCGGTGACAGGCCCGCGGTTCTTC 60
Db 1 MetLeuSerProAlaThrGlnTyrAlaAlaAlaValGlnAlaAlaProSerPheLeu 20
Qy 61 GATCTGACCGCAGCTGACGAAAGCGATCGGCTGATGACAGCGCGCCAAAGCAGAC 120
Db 21 AspLeuAspArgThrAlaGluLysThrIleAlaIleIleGluGlnAlaAlaGluGlnAsp 40
Qy 121 GTGCGCTGATGCGATTCCCAAGACTTGGAATTCGCGCTATTCCTTTGGATATGGCTG 180
Db 41 ValArgLeuIleAlaPheProGluThrPileProGlyTyrProLeuThrPileTrpLeu 60
Qy 181 GCGCGCCGCGCTTGGGAGCATGCGCTTCGCAAGCGCTATTCGAGAAATTCGCTGCGCC 240
Db 61 GlySerProAlaTrpGlyMetArgPheValGlnArgTyrPheGluAsnSerLeuValArg 80
Qy 241 GCGAGCAACAGTGGCAGGCGCTGGCGGATGCGCGCGCGCGCCGACGCGCATGATCGTG 300
Db 81 GlySerLysGlnThrPsnAlaIleAlaAspAlaAlaArgArgHisArgMetThrValVal 100
Qy 301 GCGCGCTATGAGAGCGCGCGCGCGCAAGCTCAATGAGCGCAAGCGCATTCGCGCCG 360
Db 101 ValGlyPheSerGluArgAlaGlyGlySerLeuTyrMetGlyGlnAlaIlePheGlyPro 120
Qy 361 GATGCGCATCTGATCGCGCGCGCGCAAGCTCAAGCTCAAGCTCAAGCTCAAGCTCAAG 420
Db 121 GluGlyGluLeuIleAlaAlaArgArgLysLeuLysProThrHisAlaGluThrValVal 140
Qy 421 TTCGGCAGAGAGAGACGCGCGCATCTCGCGGTGCACAGATACCGCATCGCGCGCTCGGC 480
Db 141 PheGlyGluGlyAspLysSerHisLeuAlaValTyrGluThrGlyValGlyArgIleGly 160
Qy 481 GCGCTGTGTTGGGAGACATCCAGCATTCGTAATGCAATACGCAATGATACCGCGCGAC 540
Db 161 AlaLeuCysCysTrpGluHisIleGlnProLeuSerLysTyrAlaMetTyrAlaAlaAsn 180
Qy 541 GAAACAGTCCACGTCGCGCTGCGCGCGCAAGCTTACGCGGTGAGGCGCGCTTATGCG 600
Db 181 GluGlnValHisValAlaSerTrpProCysPheSerLeuTyrArgGlyMetAlaTyrAla 200
Qy 601 CTCGACCGGAGGTCAATACCGCGCGCAAGCTTACGCGGTGAGGCGCGCTTATGCG 660
Db 201 LeuGlyProGluValAsnThrAlaAlaSerGlnValTyrAlaValGluGlyCysTyr 220
Qy 661 GTGTCGCGCTGTCGCGCAGCGCTTTCGCGGAGATGATCAAGATTTGTGTGATACGCC 720
Db 221 ValLeuAlaSerCysLeuValValThrProGluIleLeuValValLeuIleAspThrPro 240
Qy 721 GACAAAGAGATGTTCTCTCAAGCGCGCGCGGCTTTGCCATGATTTTCGGGCCGACGGC 780
Db 241 AspLysGluProLeuLeuLeuAlaGlyGlyPheSerMetIlePheGlyProAspGly 260
Qy 781 CGGCGCTGGCGGCGCTCCCGAGACCGAAGAGGAGCTGGCTGGCGCGCATATGAC 840
Db 261 ArgAlaLeuAlaGlnProLeuProGluThrGluGluLeuValThrAlaGluIleAsp 280
Qy 841 CTCGGCATGATCGCGTTGGCCCAAGCGCGCGCGATCCGCGCGCATATTCACGCGCC 900
Db 281 LeuGlyAlaIleAlaLeuAlaValAlaAlaAlaAspProAlaGlyHisTyrAlaArgPro 300
Qy 901 GAGTAAACGGCGCTGGCTGGATTCGACGTCGCGCCCAAGCGCTGGCATGATGCC 960
Db 301 AspValThrArgLeuLeuLeuAsnProArgProAlaAlaArgValGluAlaGluGlyPro 320
Qy 961 GCATTGCAACCGCAAAACGAGACAAAGGCGACGCGCC 999
Db 321 ArgPheGluValValGlnSerGluGlnAlaGluProPro 333
```

## RESULT 10

US-10-440-503-34  
; Sequence 34, Application US/10440503  
; Publication No. US20040038419A1



```

Db      21 AspleuaspArgThrValGluLysThrIleAlaIleLeuGlnAlaAlaGluGlnAsp 40
QY      121 GTGGCGCTGATCCGATTCGAGACGAGACTTGATCCCGGCTATCCCTTTGGATATGCGCTG 180
Db      41 ValArgLeuIleAlaIleAlaPheProGlnThrTrpIleProGlyLysProLeuTrpIleTrpLeu 60
QY      181 GCGCGCGCGCTTGAGGCGATGCGCTTCGTCACGAGCTATTCAGAAATTCGCTGTCGCGC 240
Db      61 GlySerProAlaTrpGlyMetArgPheValGlnArgLysPheGlyLeuAsnSerLeuValArg 80
QY      241 GCGCGCAGAGAGTGGCGAGCGCGCTTGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 300
Db      81 GlyserySerGlnTrpAsnAlaIleAlaAspAlaAlaArgAlaAlaArgAlaAlaGlyMetThrVal 100
QY      301 GCGCGCTATAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 360
Db      101 ValGlyPheSerGluArgAlaGlyGlySerLeuLysMetGlyGlnAlaIlePheGlyPro 120
QY      361 GATGGCGATCTGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 420
Db      121 GluGlyGluLeuIleAlaAlaArgAlaLysLeuLysProThrIleAlaGlnArgThrVal 140
QY      421 TTCGGCAGAGGAGA CGGCGAGCGCATCTCGCGGTGACGATACCGCGCATCGGCGCGCTCGCG 480
Db      141 PheGlyLysGluGlyAspGlySerHisLeuAlaValLysGluThrGlyValGlyArgLysGly 160
QY      481 GCGCTGTTGCTGGGAGCAGCATCCAGCGCATTCGCGAAATTCGCGCATGTCGCGCGCGCG 540
Db      161 AlaLeuLysCysTrpGluHisLysIleGlnProLeuSerLysTrpAlaMetCysTrpAlaAlaAsn 180
QY      541 GAACAGGTCCACGTCGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 600
Db      181 GluGlnValHisValAlaSerTrpProCysPheSerLeuLysTrpArgGlyMetAlaTrpAla 200
QY      601 CTGGGACCGGAGGTCATACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTG 660
Db      201 LeuGlyProGluValAlaThrAlaAlaSerGlnValLysValGluGlyCysTrp 220
QY      661 GTGCTGGCTGCTGTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 720
Db      221 ValLeuAlaSerCysLeuValAlaThrProGluIleLeuLysValLeuIleAspThrPro 240
QY      721 GACAAAGAGATGTTCTCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 780
Db      241 AspLysGlnProLeuLeuLeuAlaGlyGlyPheSerMetIlePheGlyProAspGly 260
QY      781 CCGCGCGCTGCGCGAGCGCTCCCGGAGACCGAAGAGGAGGATGCTGTCGCGCGCATATGAC 840
Db      261 ArgAlaLeuAlaGlnProLeuProGluThrGluGluGlyLeuValIleAlaGluIleAsp 280
QY      841 CTGGGCGATGATCCGCTGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 900
Db      281 LeuGlyAlaIleAlaLeuAlaLysAlaAlaAspProAlaGlyHisTrpAlaAspPro 300
QY      901 GAGGTAAACGGGCTGCTGATTCAGCTCCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 960
Db      301 AspValThrTrpGluLeuLeuLeuAsnProArgProAlaAlaArgValGluAlaLeuGlyPro 320
QY      961 GCATTGCAACCGCAAAACGAGACGAGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 999
Db      321 ArgPheGluValValGlnSerGluGlnAlaGluProPro 333

RESULT 12
US-10-146-772-16
; Sequence 16, Application US/10146772
; Publication No. US20030124698A1
; GENERAL INFORMATION:
; APPLICANT: Short, Jay
; APPLICANT: Weiner, David
; APPLICANT: Chaplin, Jennifer
; APPLICANT: Chi, Ellen
; APPLICANT: Milan, Aileen
```

```

; APPLICANT: Desantis, Grace
; APPLICANT: Madden, Mark
; APPLICANT: Burk, Mark
; TITLE OR INVENTION: Nitricases
; FILE REFERENCE: Docket No. US20030124698A1 DIV-013US
; CURRENT APPLICATION NUMBER: US/10/146,772
; CURRENT FILING DATE: 2002-05-15
; PRIOR APPLICATION NUMBER: US 60/309,006
; PRIOR FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US 60/351,336
; PRIOR FILING DATE: 2002-01-22
; PRIOR APPLICATION NUMBER: US 60/300,189
; PRIOR FILING DATE: 2001-06-21
; PRIOR APPLICATION NUMBER: US 09/751,299
; PRIOR FILING DATE: 2000-12-28
; PRIOR APPLICATION NUMBER: US 60/254,414
; PRIOR FILING DATE: 2000-12-07
; PRIOR APPLICATION NUMBER: US 60/173,609
; PRIOR FILING DATE: 1999-12-29
; NUMBER OF SEQ ID NOS: 386
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 348
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Obtained from an environmental sample
US-10-146-772-16

Alignment Scores:
Pred. No.: 1,89e-79 Length: 348
Score: 1209.00 Matches: 230
Percent Similarity: 79.3% Conservative: 42
Best Local Similarity: 67.1% Mismatches: 69
Query Match: 64.3% Gaps: 2
DB: 4

US-09-751-299-1 (1-1041) x US-10-146-772-16 (1-348)
QY      10 CCATGACGAG--TATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 66
Db      2 ProThrSerLysGlnPheArgValAlaAlaValGlnAlaAlaProValPheLeuAspLeu 21
QY      67 GACCGCAGTCGAGAAACGATCGGCTGATCGAGACGCGCGCGCGCGCGCGCGCGCGCGCG 126
Db      22 GluGlyAlaIleSerLysGlyLysSerLeuIleGluGluAlaAlaSerAsnGlyAlaLys 41
QY      127 CTGATGCGATTCGCGAGACTTGATTCGCGGCTATCCCTTTGGATATGCGTCGCGCGCGCG 186
Db      42 LeuIleAlaPheProGluThrTrpIleProGlyLysProTrpTrpIleTrpLeuAspSer 61
QY      187 CCGGCTTGAGGAGATGCGCTTCGTCACGCGCTATTCGAGAAATTCGCTGCGCGCGCGCG 246
Db      62 ProAlaTrpGlyMetArgPheValGlnArgLysPheAspAsnSerLeuMetLeuGlySer 81
QY      247 AAGCAGTGGCAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 306
Db      82 GluGlnAlaLysArgMetCysGlnAlaAlaAlaAsnAsnLysIleLysValAlaMetGly 101
QY      307 TATAGGAGCGCGCGCGCGCGCGCGCGCGCTTATGCGCGAGCGCATCTTCGCGCGCGAGCG 366
Db      102 TysSerGluArgSerGlyLysLeuLysMetGlyGlnSerIleIleAsnAspLysGly 121
QY      367 GATCTGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 426
Db      122 GluThrIlePheThrArgTrpLysLeuLysProThrHisValGluArgThrValPheGly 141
QY      427 GAGGAGACGCGAGCGCATCTCGCGGTGACGATACCGCGCGCGCGCGCGCGCGCGCGCGCG 486
Db      142 GluGlyAspGlySerHisLeuLysValMetAspThrGluIleGlyArgValGlyAlaMet 161
QY      487 TGTTCGTGGAGACACATCCAGCGCATTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 546
Db      161 TGTTCGTGGAGACACATCCAGCGCATTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG
```



```
Db 162 CyeCyvtrpGlnIleuGlnProLeuSerIytrAlaMetYserGIlnaSpGIuGln 181
Qy 547 GTCCACGTCGGCTGTCGGCCGAGCTTCAAGCTCTATCCGGGAGATGGCTTAAGCTCGGA 606
Db 182 ILehIstleAlaSerTrpProSerPheSerLeuYrAIGyAlaIleYrAlaLeuGly 201
Qy 607 CCGAGATCAATACCGCCGAGCCAGATCTACGGCTGACGAGCGGCTGACGTCG 666
Db 202 ProGluLeuAenAenAlaIleSerGlnMetYrAlaIaGluGlyGlnCyPheValIleu 221
Qy 667 GCGTCGTCGGCAGCGCTTCCCGGAGATGATCAAGGATTTGGTGAATACGCCGAGCAAG 726
Db 222 AlaProCySaIaThrValSerIySerGlnMetIleGlnMetLeuIleAaPspRArgLyS 241
Qy 727 GAGATGTTCTTCAGAGCGCGGCGGCTTTTGCATGATTTTCGGGCGCGAGCGCGCC 786
Db 242 GluProLeuLeuLeuGluGlyGlyIlePheThrMetIleYrGlyProAaPylAArgPro 261
Qy 787 CTGGCCGAGCGCTCCCGGAGAGCCGAGAGGAGTGTGTCGCGATATGACCTCGGC 846
Db 262 LeuAlaYsProLeuProGluAenGluGluGlyLeuLeuYrAlaAaPspRleuGly 281
Qy 847 ATGATCGGCTTGGCCAGAGCGCGCGCCGATCCGCGGCGCACTATTTCAAGCGCCGAGTA 906
Db 282 MetIleSerMetAlaYrAlaIaIaIaAaPspRAlaGlyIleYrAlaIaArgProAaPyl 301
Qy 907 ACGCGCTGCTGTCGATGACGTCGCGCCCAAGCGCTGTCACGCTTGAATGCGCATTC 966
Db 302 ThrArgLeuProPheAaSerAlaProAlaAaPspRValGlyYrIleAaPspRAlaSer 321
Qy 967 GAACGCAAAACGAGGAGAGAGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1026
Db 322 GlyPro--ThrGlnSerLeuYsAaPspMetGlyYsMetGlnMetGluAaGluGln 340
Qy 1027 GCCGCGCGC 1035
Db 341 LysAlaAla 343

RESULT 13
US-10-241-742-16
; Sequence 16, Application US/10241742
; Publication No. US20040002147A1
GENERAL INFORMATION:
; APPLICANT: Short, Jay
; APPLICANT: Weiner, David
; APPLICANT: Chaplin, Jennifer
; APPLICANT: Chi, Ellen
; APPLICANT: Milan, Aileen
; APPLICANT: Desantis, Grace
; APPLICANT: Madden, Mark
; TITLE OF INVENTION: Nitriases
; FILE REFERENCE: Pocket No. US20040002147A1 DIV-013US
; CURRENT APPLICATION NUMBER: US/10/241,742
; PRIOR FILING DATE: 2002-09-09
; PRIOR APPLICATION NUMBER: US/10/146,772
; PRIOR FILING DATE: 2002-05-15
; PRIOR APPLICATION NUMBER: US 60/309,006
; PRIOR FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US 60/351,336
; PRIOR FILING DATE: 2002-01-22
; PRIOR APPLICATION NUMBER: US 60/300,189
; PRIOR FILING DATE: 2001-06-21
; PRIOR APPLICATION NUMBER: US 09/751,299
; PRIOR FILING DATE: 2000-12-28
; PRIOR APPLICATION NUMBER: US 60/254,414
; PRIOR FILING DATE: 2000-12-07
; PRIOR APPLICATION NUMBER: US 60/173,609
; NUMBER OF SEQ ID NOS: 386
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 348
```

```
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE: Unknown
; OTHER INFORMATION: Obtained from an environmental sample
US-10-241-742-16
Alignment Scores:
Pred. No.: 1,89e-79 Length: 348
Score: 1209.00 Matches: 230
Percent Similarity: 79.3% Conserved: 42
Best Local Similarity: 67.1% Mismatches: 69
Query Match: 64.3% Indels: 2
Gaps: 2

US-09-751-299-1 (1-1041) x US-10-241-742-16 (1-348)
Qy 10 CCCATGACGAAG--TATCGGCGCGCGCGGTGACGCGCGCGGCTGTCGATCTC 66
Db 2 ProThrSerIySglnPheArgValAlaIaValGlnAlaIaProValPheLeuAaPspR 21
Qy 67 GACCGCACAGTCGAGAAAGCGATGCGCTGATCGACAGCGCGCGCAAGCAGACTGCGC 126
Db 22 GluGlyAlaIleSerIySgIleSerLeuIleGluGlnAlaIaIaSerAaenGlyAlaIys 41
Qy 127 CTGATCCGATTCGCGAGAGACTTGATCCCGGCTATCCCTTTGGATATGCGTGGGCGG 186
Db 42 LeuIleAaPspRProGluThrTrpIleProGlyYrProTrpTrpIleTrpLeuAaPspR 61
Qy 187 CCGGCTTGGGCGATGCGCTTGTGTCAGCGCTATTTGAGAAATTCGTCGTCGCGCGCAGC 246
Db 62 ProIaIleTrpGlyMetAlaPheValGlnArgYrPheAaPspRleuSerMetLeuGlySer 81
Qy 247 AAGCAGTGGCAGCGCCCTGGCGGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 306
Db 82 GluGlnAlaYsArgMetAaenGlnAlaIaIaAaPspRValIyYrValValMetGly 101
Qy 307 TATAGCAGCGCGCGCGCGCGCGCTCTATATGCGCGAGGATCTCGCGCGCGATGCGC 366
Db 102 TyserGlnArgSerGlyGlySerLeuYrMetGlyIleSerIleIleAaPspRlySg 121
Qy 367 GATCTGATCGCGCGCGCGCGCGCGAGCTCAAGCTTACCCATGCGGAGCGCACCGCTTGGC 426
Db 122 GluThrIlePheThrArgArgIyLeuYsProThrIleValGlnAaGlnThrValPheGly 141
Qy 427 GAGGAGACGCGACCGATCTCGCGGTGACAGATACCGCATTCGGCGCTCGCGCGCTC 486
Db 142 GluGlyAaPspRlySerIleSleCySValMetAaPspRGlulIleGlyArgValAlaMet 161
Qy 487 TGTTCGTCGAGACCATTCGAGCCATTTGCGAAATACCGCATTCGCGCGCGCGAGACAG 546
Db 162 CyeCyvtrpGlnIstleuGlnProLeuSerIytrAlaMetYserGIlnaSpGIuGln 181
Qy 547 GTCCACGTCGGCTGTCGGCCGAGCTTCAAGCTCTATCCGGGAGATGGCTTAAGCTCGGA 606
Db 182 ILehIstleAlaSerTrpProSerPheSerLeuYrArgIyAlaIaIytrAlaLeuGly 201
Qy 607 CCGAGATCAATACCGCCGAGCCAGATCTACGGCTGACGAGCGGCTGACGTCG 666
Db 202 ProGluLeuAenAenAlaIleSerGlnMetYrAlaIaGluGlyGlnCyPheValIleu 221
Qy 667 GCGTCGTCGGCAGCGCTTCCCGGAGATGATCAAGGATTTGGTGAATACGCCGAGCAAG 726
Db 222 AlaProCySaIaThrValSerIySerGlnMetIleGlnMetLeuIleAaPspRArgLyS 241
Qy 727 GAGATGTTCTTCAGAGCGCGGCGGCTTTTGCATGATTTTCGGGCGCGAGCGCGCC 786
Db 242 GluProLeuLeuLeuGluGlyGlyIlePheThrMetIleYrGlyProAaPylAArgPro 261
Qy 787 CTGGCCGAGCGCTCCCGGAGAGCCGAGAGGAGTGTGTCGCGATATGACCTCGGC 846
Db 262 LeuAlaYsProLeuProGluAenGluGluGlyLeuLeuYrAlaAaPspRleuGly 281
Qy 847 ATGATCGGCTTGGCCAGAGCGCGCGCGGATCCGCGCGCGCGCGCGCGCGCGCGAGTA 906
```

```
Db 282 MetilsermetcAlaLysAlaAlaAlaProAlaGlyLysAlaAlaArgProAlaPval 301
QY 907 ACCGGCTGCTGCTGATCGACGTCGGGCCCAACGCGCTTCACGCTTGATGCCGATTC 966
Db 302 ThirArgLeuPheAenSerAlaProAlaAenArgValGluYrLleAenProAlaSer 321
QY 967 GAACCCGAAACGAGACAAAGGCGACCGCCCGCGCTGCGGTGCGGAAAGCGCC 1026
Db 322 GlyPro---ThirGluSerLeuLysApmetGlyLysMetGlnMetGlnAlaGluGln 340
QY 1027 GCCCGCGCG 1035
Db 341 LysAlaAla 343

RESULT 14
US-10-440-523-16
; Sequence 16, Application US/10440523
; Publication No. US20040014195A1
; GENERAL INFORMATION:
; APPLICANT: Short, Jay
; APPLICANT: Weiner, David
; APPLICANT: Chaplin, Jennifer
; APPLICANT: Chi, Ellen
; APPLICANT: Milan, Aileen
; APPLICANT: Desantis, Grace
; APPLICANT: Burck, Mark
; TITLE OF INVENTION: Nitric Oxides
; FILE REFERENCE: Docket No. US20040014195A1 DIV-013US
; CURRENT APPLICATION NUMBER: US/10/440,523
; CURRENT FILING DATE: 2003-05-15
; PRIOR APPLICATION NUMBER: US/10/146,772
; PRIOR FILING DATE: 2002-05-15
; PRIOR APPLICATION NUMBER: US 60/309,006
; PRIOR FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US 60/351,336
; PRIOR FILING DATE: 2002-01-22
; PRIOR APPLICATION NUMBER: US 60/300,189
; PRIOR FILING DATE: 2001-06-21
; PRIOR APPLICATION NUMBER: US 09/751,299
; PRIOR FILING DATE: 2000-12-28
; PRIOR APPLICATION NUMBER: US 60/254,414
; PRIOR FILING DATE: 2000-12-07
; PRIOR APPLICATION NUMBER: US 60/173,609
; PRIOR FILING DATE: 1999-12-29
; NUMBER OF SEQ. ID NOS: 386
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 348
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Obtained from an environmental sample
US-10-440-523-16

Alignment Scores:
Pred. No.: 1.89e-79 Length: 348
Score: 1209.00 Matches: 230
Percent Similarity: 79.3% Conservative: 42
Best Local Similarity: 67.1% Mismatches: 69
Query Match: 64.3% Indels: 2
Db: 4 Gaps: 2

US-09-751-299-1 (1-1041) x US-10-440-523-16 (1-348)
QY 10 CCATGACGAG--TATCGCGCGCGGTGCGAGCGCGCGGTGTTCTTGATCTC 66
Db 2 ProThSerLysGlnPheArgValAlaAlaValGlnAlaAlaProValPheLeuApm 21
QY 67 GACCGCACACTCGAAGAAAGGATGCGCTGATCGAGCAGCGCGCAAGCAGACGTGCG 126
Db 22 GluGlyAlaLleSerLysGlyLleSerLeuLleGluGlnAlaAlaSerAenGlyAlaLys 41
```

```
QY 127 CTGATCGCATTCACAGACTTGATTCGCGGCTATCCCTTTGGATATGCTGGCGC 186
Db 42 LeuLleAlaPheProGluThrTrpLleProGlyLysProTrpTrpLleTrpLeuApmSer 61
QY 187 CCGGCTTGGGAGCATGCGCTTCGTCAGCGCTATTTGAGAAATTCGCTGCGCGCGAGC 246
Db 62 ProAlaTrpPolymetArgPheValGlnArgTrpPheApmSerLeuMetLeuLysSer 81
QY 247 AAGCAGTGGCAGGCGCTGGCGGATGCGCGCCCGCCGCAAGCATGATGTCGGCGCG 306
Db 82 GluGlnAlaLysArgMetAenGlnAlaAlaAlaApmAenLysLleTrpValValMetGly 101
QY 307 TATAGCAGGCGCGCGCGCGCGCTTATATAGCGCGCATTCCTGGCGCCGAGCGC 366
Db 102 TyrSerGluArgSerGlyLysLeuTrpMetGlyGlnSerLleLleApmApmLysGly 121
QY 367 GATCTGATCGCGCGCGCGCGCATCAAGCTCAAGCTCAAGCTGCGAGCGCACCGTTCGCG 426
Db 122 GluThrLlePheThrArgValGlyLysLeuLysProThrLleValGluArgThrValPheGly 141
QY 427 GAGGAGACCGCGACCGCATTCGCGGTGCAAGATACCGCCATCGCGCGCTCGCGCGCTC 486
Db 142 GluGlyApmGlySerLysLysLeuCyValMetApmThrGluLleGlyArgValGlyAlaMet 161
QY 487 TGTTCGTGGAGACATCCAGCATTCGAAATACGCGCATGACGCGCGCGCGAGCAAG 546
Db 162 CysCysTrpGluLysLeuGlnProLysSerLysTrpAlaMetLysSerGlnApmGluGln 181
QY 547 GTCCAGCTCGCGGTGCGCGCGCTTCAGCTTCATTCGCGCATGCGCTTATGCGCTTGGGA 606
Db 182 LLeHisLleAlaSerTrpProSerPheSerLeuTrpArgLysAlaAlaTrpAlaLeuGly 201
QY 607 CCGGAGGTCAATACCGCGCGCAAGCATCAAGCGGTGCGAGCGCGCTCTGCTGCTG 666
Db 202 ProGluLeuApmApmAlaAlaSerGlnMetTrpAlaAlaGluGlnCysPheValLeu 221
QY 667 GCGTCGCGCGCGCGCTTGGCGCGAGATATCAAGTATGATGATACCGCGCGAG 726
Db 222 AlaProCysAlaThrValSerLysGluMetLleGluLeuLleApmApmProAlaGlyLys 241
QY 727 GAGATGTTCTCAAGCGCGCGCGGTGTTGACATGTTTCGGGCGCGCGCGCGCGC 786
Db 242 GluProLeuLeuLeuGluGlyGlyPheThrMetLleTrpLleProApmGlyArgPro 261
QY 787 CTGCGCGAGCGCGCTCCGCGAGACCGAAGAGGACTGCTGTCGCCGATATCGACTTGGC 846
Db 262 LeuAlaLysProLeuProGluAenGluGluGlyLeuLeuTrpAlaApmLleApmLeuGly 281
QY 847 ATGATCGCTTGGCAAGCGCGCGCGCGATCCGCGCGGCGCACTATTCACGCGCGAGTA 906
Db 282 MetLleSerMetcAlaLysAlaAlaAlaApmProAlaGlyLysTrpAlaApmProApmVal 301
QY 907 ACCGGCTGCTGCTGATCGACGTCGGGCCCAACGCGCTTCACGCTTGATGCCGATTC 966
Db 302 ThirArgLeuPheAenSerAlaProAlaAenArgValGluYrLleAenProAlaSer 321
QY 967 GAACCCGAAACGAGACAAAGGCGACCGCGCGCTGCGGTGCGGAAAGCGCC 1026
Db 322 GlyPro---ThirGluSerLeuLysApmetGlyLysMetGlnMetGlnAlaGluGln 340
QY 1027 GCCCGCGCG 1035
Db 341 LysAlaAla 343

RESULT 15
US-10-440-503-16
; Sequence 16, Application US/10440503
; Publication No. US20040038419A1
; GENERAL INFORMATION:
; APPLICANT: Weiner, David Paul
; APPLICANT: Chaplin, Jennifer Ann
; APPLICANT: Chi, Ellen
```

```

: APPLICANT: Milan, Aileen
: APPLICANT: Desantis, Grace
: APPLICANT: Burk, Mark J.
: APPLICANT: McQuaid, Jeffrey
: APPLICANT: Steege, Justin
: TITLE OF INVENTION: ASSAYS AND KITS FOR DETECTING THE PRESENCE OF
: TITLE OF INVENTION: NITRILES AND/OR CYANIDE
: FILE REFERENCE: 09010-900001
: CURRENT APPLICATION NUMBER: US/10/440,503
: CURRENT FILING DATE: 2003-05-15
: PRIOR APPLICATION NUMBER: US 60/380,737
: PRIOR FILING DATE: 2002-05-15
: NUMBER OF SEQ ID NOS: 386
: SOFTWARE: FastSeq for Windows Version 4.0
: SEQ ID NO: 16
: LENGTH: 348
: TYPE: PRT
: ORGANISM: Unknown
: FEATURE:
: OTHER INFORMATION: Obtained from an environmental sample
: US-10-440-503-16

Alignment Scores:
Pred. No.: 1,896-79 Length: 348
Score: 1209.00 Matches: 230
Percent Similarity: 79.3% Conservative: 42
Best Local Similarity: 67.1% Mismatches: 69
Query Match: 64.3% Indels: 2
DB: 4 Gaps: 2

US-09-751-299-1 (1-1041) x US-10-440-503-16 (1-348)
QY 10 CCCATGACGAG--TATCGCGCGCGCGGTGACGCGCGCGGTTCCTGATCTC 66
DB 2 ProThrSerLysGlnPheArgValAlaAlaValGlnAlaAlaProValPheLeuAspLeu 21
QY 67 GACCCGACAGTCGAGAAAGCCATCGCGCTGATCGACGCGCGCCCAAGCAGAGACTGCGC 126
DB 22 GlnGlyAlaIleSerLysGlyIleSerLeuIleGlnGlnAlaIleSerAsnGlyAlaLys 41
QY 127 CTGATCGATTCCTCCAGAGACTGATTCCTCCGCTATCCCTTTTGATATGCTGCGCGCG 186
DB 42 LeuIleAlaPheProGlnThrTrpIleProGlyTyrProTrpTrpIleTrpLeuAspSer 61
QY 187 CCGGCTTGGGCGATCGCTTGTCTGACGCGCTATTTGAGAAATTCGTCGTGCGCGCGAC 246
DB 62 ProAlaTrpGlyMetArgPheValGlnArgTyrPheAspAsnSerLeuMetLeuGlySer 81
QY 247 AAGCGTGGCAGCGCCCTGCGGATCGCGCCGCGCGCGCGCATGATGCTGCGCGCGC 306
DB 82 GlnGlnAlaLysArgMetAsnGlnAlaAlaAlaAsnAsnLysIleTyrValValMetGly 101
QY 307 TATAGCAGCGCGCGCGCGCGAGCTCTATATATGCGCGAGCGATCTTCGCGCCGATGCG 366
DB 102 TyrSerGlnArgSerGlyIleSerLeuTyrMetGlyGlnSerIleIleAsnAspLysGly 121
QY 367 GATCTGATCGCGCGCGCGCGCAAGCTCAAGCTTACCTACCATGCGGAGCGCACCGTTCG 426
DB 122 GlnTrpIlePheThrArgArgLysLeuLysProThrIleValGlnArgThrValPheGly 141
QY 427 GAGGAGACGCGCGAGCATCTCGCGGTGACAGATACCGGCATCGGCGCGCTCGCGCGCTC 486
DB 142 GlnGlyAspGlySerHisLeuLysValMetAspThrGlnIleGlyArgValGlyAlaMet 161
QY 487 TGTTCCTGGAGACATCAAGCATTTGTGAATAAGCATATGATGACGCGCGAGCAAGAG 546
DB 162 CysCysTrpGlnHisLeuGlnProLeuSerLysTyrAlaMetTyrSerGlnAspGlnGln 181
QY 547 GTCCAGTCGCGCTGCGCGCGAGCTTCAAGCTTATCGCGGCGATGAGCTATGCGCTCG 606
DB 182 IleHisIleAlaSerTrpProSerPheSerLeuTyrArgGlyAlaAlaIleTyrAlaLeuGly 201
QY 607 CCGAGAGTCATACCGCGCGCAAGCCAGATCTACGCGGTGAGGCGCGGCTGCTAGTGTG 666
```

```

DB 202 ProGlnLeuAsnAlaAlaIleSerGlnMetTyrAlaAlaGlnGlyGlnCysPheValLeu 221
QY 667 GCGTCGCGCGAGCGTTTTCGCGCGAGATGATCAAGTATTTGATGATACGCCGAGACAG 726
DB 222 AlaProCysAlaThrValSerLysGlnMetIleGlnMetLeuIleAspProArgLys 241
QY 727 GAGATGTTCTCAAGCGCGCGGCTTTCGATGATTTTTCGCGCGCGCGCGCGCGCGC 786
DB 242 GlnProLeuLeuGlnGlyGlyGlyPheThrMetIleTyrGlyProAspGlyArgPro 261
QY 787 CTGCGCGAGCGCTCCGCGAGACCGAAGAGGACTGCTGTCGCGCATATGACCTCGGC 846
DB 262 LeuAlaLysProLeuProGlnAsnGlnGlyIleLeuLeuTyrAlaAspIleAspLeuGly 281
QY 847 ATGATCGGTTGGCAAGCGCGCGCGCGATCCGCGCGCGCGCACTTATTCAGCGCGCG 906
DB 282 MetIleSerMetAlaLysAlaAlaAlaAspProAlaGlyHisTyrAlaArgProAspVal 301
QY 907 ACGCGCTGCTGCTGATTCGACGTCGCCCGCAAGCGCTTCGACGCTTATGCGCGCATTC 966
DB 302 ThrArgLeuLeuPheAsnSerAlaProAlaAsnArgValGlnTyrIleAsnProAlaSer 321
QY 967 GAACCGAATAACGAGACAGAGCGCGCGCGCGCGCTGCGCGGTGCGCGAAGCGCC 1026
DB 322 GlyPro--ThrGlnSerLeuLysAspMetGlyLysMetGlnMetGlnAlaGlnGln 340
QY 1027 GCCGCGCGCG 1035
DB 341 LysAlaAla 343
```

Search completed: April 27, 2006, 01:51:41  
Job time : 172.128 secs

**THIS PAGE BLANK (USPTO)**



```
QY 25 GCGGCGCGCGGTGACAGCGCGCGGTGTTCTCGATTCGACCGCAGTCGAGAA 84
Db 8 ArgAlaAlaAlaValGlnAlaAlaSerProAsnYrLeuAlaAlaThrGlyValAlaAsp 27
QY 85 GCGATCGCGCTGATCGAGCGCGCGCAGAGCGAGCGCGCGCTGATTCGATTCGACGAG 144
Db 28 ThrIleGluLeuAlaArgGlnAlaArgAspGluGlyCysAspLeuIleValPheGlyGlu 47
QY 145 ACTTGATTCGCGCTATCCCTTTTGGATATGCTGGGCGCGCGCGCTTGGGCGCATGCGC 204
Db 48 ThrTrpLeuProGlyTyrProPheHisValTrpLeuGlyAlaProAlaTrpSerLeu 67
QY 205 TTTCGTCAGCGCTATTTGAGAAATTCGCTCGTGGCGCGCGCAGACAGCGTGGCAGCCCTG 264
Db 68 TyrsSerAlaArgTyrTyrAlaAsnSerLeuAspSerAlaGluPheGlnArgIle 87
QY 265 GCGGATCGCGCGCGCGCGCAGCGCATGTCATGTCGCGCGCTATAGCAGCGCGCGCGC 324
Db 88 AlaGlnAlaAlaArgThrLeuGlyIlePheIleAlaLeuGlyTyrSerGluArgSerGly 107
QY 325 GCGAGCTCTATATGCGCGCAGCGCGCTTTCGCGCGCGCGCGCTATGATTCGCGCGCGC 384
Db 108 GlySerLeuTyrLeuGlyGlnCysLeuLeuAspAspGlyGluLeuLeuTrpSerArg 127
QY 385 CGCAAGCTCAAGCTTACCGATGCGGAGCGCAGCGCTTTCGCGCGAGGAGAGCAGCGACCAT 444
Db 128 ArgIysLeuLysProThrHisValGlnArgThrValPheGlyGluGlyTyrAlaArgAsp 147
QY 445 CTCGCGGTGACAGTACCGCGCGCGCGCTGCGCGCGCTCTGTCGTCGAGCAGCATC 504
Db 148 LeuIleValSerAspThrGlnLeuGlyAlaValGlyAlaLeuCysGlyTrpGluHisLeu 167
QY 505 CAGCATTTGCGAAATAGCCCATGTAGCGCGCGCGCAGAACAGTCCAGTCGTCGTGG 564
Db 168 SerProLeuSerLysTyrAlaLeuTyrSerGlnHisGlnAlaIleHisIleAlaAlaTrp 187
QY 565 CCGAGCTTACGCTCTATCCGCGCATGCGCATGCGCTGCGCGCGCAGAGGTCATACCGCC 624
Db 188 ProSerPheSerLeuTyrSerGlnGlnAlaHisIleAlaLeuSerAlaLysValAsnMetAla 207
QY 625 GCAAGCCAGATCTACGCGGTGAGAGCGCGCTCTACGTCGTCGCGTCGTCGCGCTT 684
Db 208 AlaSerGlnIleTyrSerValGlnGlyGlnCysPheThrIleAlaAlaSerSerVal 227
QY 685 TCGCGCGATGATGATCAAGATTTGTCGATACGCGCGCAGAGAGATGTTCTCAAGGCC 744
Db 228 ThrGlnGluThrIleAspMetLeuGluValGlyGluHisAsnAlaProLeuLeuLysVal 247
QY 745 GCGCGCGCTTTTGCAGATTTTCGAGCGCGCAGCGCGCGCTGCGCGCAGCGCTCCG 804
Db 248 GlyIleGlySerSerMetIlePheAlaProAspGlyArgThrLeuAlaProTyrLeuPro 267
QY 805 GAGACGAGAGGAGTCTGTCGTCGCGATACGCTCGGATGATCGGCTTGGCGCAAG 864
Db 268 HisAspAlaGlnGlyLeuIleIleAlaAspLeuAsnMetGluGluIleAlaPheAlaLys 287
QY 865 GCGCGCGCGATCCGCGCGCGCATATTCACGCGCAGCATACCGCGCTGCTGTCGAT 924
Db 288 AlaIleAsnAspProValGlyHisTyrSerLysProGluAlaThrArgLeuValLeuAsp 307
QY 925 -----CGAGCTCGCGCGCAGCGCGTCTGACGCTTATGCGCATTCGACGAGCGCA 975
Db 308 LeuGlyHisArgAspPrometThrArgVal-----HisSerLys 320
QY 976 AAGGAGGACAAAGCGCGCGCGCGCTGCGCGTGTGCGGAGAACCGCGCGCGCG 1035
Db 321 SerValThrArgGlnGluAlaProGluGlnGlyValGlnSerLysIleAlaSerAla 340
```

RESULT 2  
US-10-919-182-16  
; Sequence 16, Application US/10919182  
; Publication No. US2006003532A1

```
GENERAL INFORMATION:  
; APPLICANT: E.I. duPont de Nemours and Company, Inc.  
; APPLICANT: Di Cosimo, Robert  
; APPLICANT: Payne, Mark  
; APPLICANT: O'Keefe, Daniel  
; TITLE OR INVENTION: PRODUCTION OF 3-HYDROXYCARBOXYLIC ACID USING NITRILASE MUTANTS  
; FILE REFERENCE: C12584 US NA  
; CURRENT APPLICATION NUMBER: US/10/919,182  
; NUMBER OF SEQ ID NOS: 18  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 16  
; LENGTH: 369  
; TYPE: PRT  
; ORGANISM: Artificial sequence  
; FEATURE:  
; OTHER INFORMATION: Mutant nitrilase with Phe168 to Val change  
US-10-919-182-16  
Alignment Scores:  
Pred. No.: 1 536-48 Length: 369  
Score: 810.00 Matches: 160  
Percent Similarity: 62.1% Conservative: 53  
Best Local Similarity: 46.6% Mismatches: 126  
Query Match: 43.1% Indels: 4  
DB: 6 Gaps: 2  
US-09-751-299-1 (1-1041) x US-10-919-182-16 (1-369)  
QY 16 ACAGAAATATCGCGCGCGCGCGCGCGCGCGCTTCTCGATTCGACCGCAGCA 75  
Db 6 SerIysLeuAlaAlaThrValGlnAlaGluProValTrpLeuAspAlaSerAlaTrp 25  
QY 76 GTGAGAAAGCATCGCGCTGATCGAGCAGCGCGCAGACAGACGTCGCGCTGATCGCA 135  
Db 26 IleAspLysSerIleGlyIleIleGluGlnAlaGlnLysGlyAlaSerLeuIleAla 45  
QY 136 TTTCCAGACACTTGATTTCCCGGCTATCCCTTTGGATATGCTGGCGCGCGCGCTTGG 195  
Db 46 PheProGluValPheIleProGlyTyrProTyrTrpAlaTrpLeuLysAspValLysTyr 65  
QY 196 GGCATCGCTTCTGTCGCGCTATTTGAGAAATTCGCTGTCGCGCGCAGCAAGCATGG 255  
Db 66 SerLeuSerPheThrSerArgTyrHisGluAsnSerLeuGlnLeuGlyAspAspArgMet 85  
QY 256 CAGGCGCTGCGGATGCGCGCGCGCGCGCAGCATGCTGTCGCGCGCTATACCGAG 315  
Db 86 ArgArgLeuGlnLeuAlaAlaArgAsnLysIleAlaLeuValMetLysTyrSerGlu 105  
QY 316 GCGCGCGGCGGAGCGCTCTATATGAGCGCAGCGCATCTTCGCGCGATGCGATGATC 375  
Db 106 ArgGluAlaGlySerArgTyrLeuSerGlnValPheIleAspGluArgGlyGluIleVal 125  
QY 376 GCGCGCGCGCGAGGCTCAAGCTCAAGCTCAAGCTGCGAGCGCAGCGCTTGGCGAGAC 435  
Db 126 AlaAsnArgArgLysLeuLysProThrHisValGlnArgThrIleTyrGlyGluLysAsn 145  
QY 435 GCGAGCATCTGCGGATGACAGTACCGGCATCGGAGCGCGCTTGGCGCGCTGTCGCG 495  
Db 146 GlyThrAspPheLeuThrHisAspPheAlaPheGlyArgValGlyIleuAsnCysTrp 165  
QY 496 GAGCAGATCAACGATTTGCGAAATACGCAATGTCAGCGCGCGCAGAGCATGTCACGTC 555  
Db 166 GluHisValGlnProLeuSerLysPheMetCysTyrSerLeuGlyGlnGlnValHisVal 185  
QY 556 GCGTCGTGCGCGAGCTTACGCTTATCGCGCGCATGCGCTATGCGCTCGACCGAGAGTC 615  
Db 186 AlaSerTrpProAlaMetSerProLeuGlnProAspValPheGlnLeuSerIleGluAla 205  
QY 616 AATACGCGCGCAGACGATACGCGGATCGAGCGCGCGCTGTCAGTCTGCGCGCTGCGC 675  
Db 206 AsnAlaTrpValThrArgSerTyrAlaIleGluGlyGlnThrPheValLeuCysSerThr 225
```

Qy 676 GCGACGTTTCCGCGAGATGATCAAGGATTTGGATGATACCGCGACAGAGATGTTTC 735  
Db 226 GlnValIleGlyProSerAlaIleGluThrPheCysLeuAsnAspGluGlnArgAlaLeu 245  
Qy 736 CTCAAGCGCGCGCGGCTTTTTCATGATTTTCGGGCGGACGCGCGCGCTTGCGCGAG 795  
Db 246 LeuProGlnGlyCysGlyTyrPheAlaArgIleTyrGlyProAspGlySerGluLeuAlaLys 265  
Qy 796 CCGCTCCGCGAGACCGAAGAGGAGTCTGTGTCGCCGATATGACCTTGCGCATGTCGCG 855  
Db 266 ProLeuAlaGluAspAlaGluGlyLeuTyrAlaGluIleAspLeuGluGlnIleLeu 285  
Qy 856 TTGGCCAGAGCGCGCGCGATCCGCGGCGCATTTATCAAGCGCGCGCATCAAGCGCGCTG 915  
Db 286 LeuAlaLysAlaGlyAlaAspProValGlyHisTyrSerArgProAspValLeuSerVal 305  
Qy 916 CTGCTGATCGACGT-----CCGCGCCACCGCGTCTGTCACGCTTGATCCGCATTC 966  
Db 306 GlnPheAspProArgAsnHisThrProValHisArg---IleGlyIleAspGlyArgLeu 324  
Qy 967 GAACCGCAAAACGAGACAGAGGCGCGCGCGCTGCGGTGCTGCGGAAAGCGCC 1026  
Db 325 AspValaenThrArgSerArgValGluAsnPheArgLeuArgGlnAlaIleGluGlnGlu 344  
Qy 1027 GCCGCGCGCG 1035  
Db 345 ArgGlnAla 347

RESULT 3  
US-10-919-182-18  
; Sequence 18, Application US/10919182  
; Publication No. US20060035352A1

; GENERAL INFORMATION:  
; APPLICANT: E.I. duPont de Nemours and Company, Inc.  
; APPLICANT: Di Cosimo, Robert  
; APPLICANT: Payne, Mark  
; APPLICANT: O'Keefe, Daniel  
; TITLE OF INVENTION: PRODUCTION OF 3-HYDROXYCARBOXYLIC ACID USING NITRILASE MUTANTS  
; FILE REFERENCE: CL2584 US NA  
; CURRENT APPLICATION NUMBER: US/10/919,182  
; NUMBER OF SEQ ID NOS: 18  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO 18  
; LENGTH: 369  
; TYPE: PRT  
; ORGANISM: Artificial sequence  
; FEATURE:  
; OTHER INFORMATION: Mutant nitrilase with Phe168 to Leu change  
US-10-919-182-18

Alignment Scores:  
Pred. No.: 1.8e-48 Length: 369  
Score: 809.00 Matches: 160  
Percent Similarity: 62.1% Conservative: 53  
Best Local Similarity: 46.68 Mismatches: 126  
Query Match: 43.0% Indels: 4  
DB: 6 Gaps: 2

US-09-751-299-1 (1-1041) x US-10-919-182-18 (1-369)

Qy 16 ACGAAGTATCCGCA 75  
Db 6 SerLysPheLeuAlaIleThrValGlnAlaGluProValITrPLeuAspAlaAspAlaThr 25  
Qy 76 GTCCGAAGAGCGATCGCGCTGATCGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCA 135  
Db 26 IleAspLysSerIleGlyIleIleGluGluAlaIleGlnLysGlyAlaSerLeuIleAla 45  
Qy 136 TTCCCGAGACTTGATTTCCGCGCTATCCCTTTTGATGATGCGTGGCGCGCGCGCTTG 195  
Db 46 PheProGluValPheIleProGlyTyrProTyrTrpAlaIlePheGlyAspValLysTyr 65

Qy 196 GGCATGCGCTTCGTCACAGCTATTTCCAGAAATTCGCTCGGCGCGCGCGCGCGAGCATGTG 255  
Db 66 SerLeuSerPheThrSerArgTyrHisGluAsnSerLeuGluLeuGlyAspArgMet 85  
Qy 256 CAGGCGCTTGCGGATGCG 315  
Db 86 ArgTrpLeuGlnLeuAlaIleArgAsnLysIleAlaLeuValMetGlyTyrSerGlu 105  
Qy 316 CGCGCGCGCGCGCGCGCTTATATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCATC 375  
Db 106 ArgGluAlaGlySerArgTyrLeuSerGlnValPheIleAspGluArgGlyGluIleVal 125  
Qy 376 GCCG 435  
Db 126 AlaAsnArgGlyLysLeuLysProThrHisValGluGlnGlnIleTyrGlyGluGlyAsn 145  
Qy 436 GCGACCATTCG 495  
Db 146 GlyThrAspPheLeuThrHisAspPheAlaPheGlyArgValGlyGlyLeuAsnCysTrp 165  
Qy 496 GAGCAGATCCAGCGCATTTGTCAATACGCCATGTACGCCCGCGCGCGCGCGCGCGCG 555  
Db 166 GluHisLeuGlnProLeuSerLysPheMetTyrSerLeuGlyGluGlnValHisVal 185  
Qy 556 GCGTGGCG 615  
Db 186 AlaSerTrpProAlaMetSerProLeuGlnProAspValPheGlnLeuSerIleGluAla 205  
Qy 616 AATACCG 675  
Db 206 AsnAlaThrValThrArgSerTyrAlaIleGluGlyGlnThrPheValLeuCysSerThr 225  
Qy 676 GCGACGTTTCCGCGGAGATGATCAAGGTATTTGGATGATGCGCGCGCGCGCGCGCG 735  
Db 226 GlnValIleGlyProSerAlaIleGluThrPheCysLeuAsnAspGluGlnAlaLeu 245  
Qy 736 CTCAAGCGCGCGCGGTTTTCATGATTTTCGGGCGGCGCGCGCGCGCGCGCGCGAG 795  
Db 246 LeuProGlnGlyCysGlyTyrPheAlaArgIleTyrGlyProAspGlySerGluLeuAlaLys 265  
Qy 796 CCGCTCCGCGAGACCGAAGAGGAGTCTGTGTCGCCGATATGACCTTGCGCATGTCGCG 855  
Db 266 ProLeuAlaGluAspAlaGluGlyIleLeuTyrAlaGluIleAspLeuGluGlnIleLeu 285  
Qy 856 TTGGCCAGAGCGCGCGCGCGATCCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 915  
Db 286 LeuAlaLysAlaGlyAlaAspProValGlyHisTyrSerArgProAspValLeuSerVal 305  
Qy 916 CTGCTGATCGACGT-----CCGCGCCACCGCGTCTGTCACGCTTGATCCGCATTC 966  
Db 306 GlnPheAspProArgAsnHisThrProValHisArg---IleGlyIleAspGlyArgLeu 324  
Qy 967 GAACCGCAAAACGAGACAGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1026  
Db 325 AspValaenThrArgSerArgValGluAsnPheArgLeuArgGlnAlaIleGluGlnGlu 344  
Qy 1027 GCCGCGCGCG 1035  
Db 345 ArgGlnAla 347

RESULT 4  
US-10-919-182-4  
; Sequence 4, Application US/10919182  
; Publication No. US20060035352A1  
; GENERAL INFORMATION:  
; APPLICANT: E.I. duPont de Nemours and Company, Inc.  
; APPLICANT: Di Cosimo, Robert  
; APPLICANT: Payne, Mark  
; APPLICANT: O'Keefe, Daniel  
; TITLE OF INVENTION: PRODUCTION OF 3-HYDROXYCARBOXYLIC ACID USING NITRILASE MUTANTS  
; FILE REFERENCE: CL2584 US NA  
; CURRENT APPLICATION NUMBER: US/10/919,182  
; CURRENT FILING DATE: 2004-08-16

```

:      NUMBER OF SEQ. ID NOS: 18
:      SOFTWARE: PatentIn version 3.2
:      SEQ ID NO 4
:      LENGTH: 369
:      TYPE: PRT
:      ORGANISM: Acidovorax facilis 72W
US-10-919-182-4

Alignment Scores:
Pred. NO.:      2,47e-48      Length:      369
Score:          807.00      Matches:      160
Percent Similarity: 61.8%      Conservative: 52
Best Local Similarity: 46.6%      Mismatches:  127
Query Match:    42.9%      Indels:      4
DB:             6      Gaps:        2
US-09-751-299-1 (1-1041) x US-10-919-182-4 (1-369)

```



```

RESULT 6
US-10-919-182-12
; Sequence 12, Application US/10919182
; Publication No. US20060035352A1
; GENERAL INFORMATION:
; APPLICANT: E.I. duPont de Nemours and Company, Inc.
; APPLICANT: Di Cosimo, Robert
; APPLICANT: Payne, Mark
; APPLICANT: O'Keefe, Daniel
; TITLE OF INVENTION: PRODUCTION OF 3-HYDROXYCARBOXYLIC ACID USING NITRILASE MUTANTS
; FILE REFERENCE: C12584 US NA
; CURRENT APPLICATION NUMBER: US/10/919,182
; CURRENT FILING DATE: 2004-08-16
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 12
; LENGTH: 369
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Mutant nitrilase with Thr210 to Cys change
US-10-919-182-12
Alignment Scores:

```

[illegible]

```
QY 967 GAACCGCAAAACGAGCAAGGGCGCGCGCTGCGCGTGTGCGGAAAGGCC 1026
Db 325 AspAlaAsnThrArgSerArgValGluAsnPheArgLeuArgGlnAlaAlaGluGlnGlu 344
QY 1027 GCCCGCCGCG 1035
Db 345 ArgGlnAla 347

RESULT 7
US-10-919-182-14
; Sequence 14, Application US/10919182
; Publication No. US2006003532A1
; GENERAL INFORMATION:
; APPLICANT: E.I. duPont de Nemours and Company, Inc.
; APPLICANT: Di Cosimo, Robert
; APPLICANT: Payne, Mark
; APPLICANT: O'Keefe, Daniel
; TITLE OF INVENTION: PRODUCTION OF 3-HYDROXYCARBOXYLIC ACID USING NITRILASE MUTANTS
; FILE REFERENCE: CL2584 US NA
; CURRENT APPLICATION NUMBER: US/10/919,182
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patent version 3.2
; SEQ ID NO 14
; LENGTH: 369
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Mutant nitrilase with Phe168 to Lys change
US-10-919-182-14

Alignment Scores:
Pred. No.: 3.98e-48 Length: 369
Score: 804.00 Matches: 160
Percent Similarity: 61.8% Conservative: 52
Best Local Similarity: 46.6% Mismatches: 127
Query Match: 42.7% Indels: 4
DB: 6 Gaps: 2

US-09-751-299-1 (1-1041) x US-10-919-182-14 (1-369)

QY 16 ACGAAGTATCGCGCGCGCGCGTGCAGCGCGCGGTGTTCTTCATTCGACCGCACA 75
Db 6 SerLysPheLeuAlaAlaThrValGlnAlaGluProValThrPheLeuAlaAspAlaThr 25
QY 76 GTCGAGAAACGATCGCGCTGATGAGCAGCGCGCGCAAGAGCAGCGCGCTGATGCA 135
Db 26 IleAspLysSerIleGlyIleIleGluGlnAlaAlaGlnLysGlyAlaSerLeuIleAla 45
QY 136 TTCGACAGACTTGATTCGCGCTATCCCTTTGATGATGCTGGCGCGCGCTTGG 195
Db 46 PheProGluValPheIleProGlyTyrProTyrThrPheIlePheGluValLysTyr 65
QY 196 GGCATGCGCTTCGCGCGCGCTATTTGAGAAATTCGTCGCGCGCGCAGACGACTGG 255
Db 66 SerLeuSerPheThrSerArgTyrLysGlnSerLeuLysGlnLysPheArgMet 85
QY 256 CAGGACCTGCGGATGCGCGCGCGCGCGCGCGCGATGATGCTGCGCGCTAAGGAG 315
Db 86 ArgArgLeuGlnLeuAlaAlaArgArgAlaLysIleLeuValMetGlyTyrSerGln 105
QY 316 CGGCGCGCGCGCGCGCTTATATGAGCGCGCGCGATCTTCGCGCGCGCGATCTGANC 375
Db 106 ArgGlnAlaGlySerArgTyrLeuSerGlnValPheIleAspGlnArgGlyGlnIleVal 125
QY 376 GCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAG 435
Db 126 AlaAsnThrArgLysLeuLysProLysValAlaGlnArgThrIleTyrGlyGlnGlyAsn 145
QY 436 GGCAGCGCATCTCGCGGTGCGAGATACCGCATCGCGCGCGCGCTGTGTGCTGG 495
Db 146 GlyThrAspPheLeuThrIleAspPheAlaPheGlyArgValGlyGlyLeuAsnCysTrp 165
```

```
QY 496 GAGCAATCATCGCATTTGTCGAATAAGCATATGACCGCGCGCGCAAGACGATCTCACGTC 555
Db 166 GlnIleLysPheLysPheLeuSerLysPheMetMetTyrSerLeuGlyGlnValIleIleVal 185
QY 556 GCGTCGTGCGCGCGCGCTTACGCGCTTATCGCGCGCGCGCGCGCGCGCGCGCGAGTC 615
Db 186 AlaSerTrpProAlaMetSerProLeuGlnProAspValPheGlnLeuSerIleGlnAla 205
QY 616 AATACCGCGCGCAAGCAGATCTACGCGGTGAGAGCGCGCGCTGACGTCGTGCGTCGCG 675
Db 206 AsnAlaThrValThrArgSerTyrAlaIleGlnGlyGlnThrPheValLeuCysSerThr 225
QY 676 GCGACGCTTTCGCGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 735
Db 226 GlnValIleLysProSerAlaIleGlnThrPheCysLeuAsnAspGluGlnArgAlaLeu 245
QY 736 CTCGAGCGCGCGCGCGCTTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAG 795
Db 246 LeuProGlnGlyCysGlyTyrPheAlaArgIleTyrGlyProAspGlySerGlnLeuAlaLys 265
QY 796 CCGCTCCCGGAGACCGAAGGAGACTGCTGCTGCTGCGCGATGATGATGATGATGATGCGG 855
Db 266 ProLeuAlaGlnAspAlaGlnGlyIleLeuTyrAlaGlnIleAspLeuGlnIleLeu 285
QY 856 TTGCGCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTG 915
Db 286 LeuAlaLysAlaGlyAlaAspProValGlyHisTyrSerArgProAspValLeuSerVal 305
QY 916 CTGCTGATCGACGT-----CCGCGCGCAAGCGGTGTCACGCTTGATCCGCGATTTC 966
Db 306 GlnPheAspProArgSerHisThrProValHisArg---IleGlyLeuAspGlyArgLeu 324
QY 967 GAACCGCAAAACGAGCAAGGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1026
Db 325 AspValAsnThrArgSerArgValGluAsnPheArgLeuArgGlnAlaAlaGluGlnGlu 344
QY 1027 GCCCGCCGCG 1035
Db 345 ArgGlnAla 347

RESULT 8
US-10-919-182-8
; Sequence 8, Application US/10919182
; Publication No. US2006003532A1
; GENERAL INFORMATION:
; APPLICANT: E.I. duPont de Nemours and Company, Inc.
; APPLICANT: Di Cosimo, Robert
; APPLICANT: Payne, Mark
; APPLICANT: O'Keefe, Daniel
; TITLE OF INVENTION: PRODUCTION OF 3-HYDROXYCARBOXYLIC ACID USING NITRILASE MUTANTS
; FILE REFERENCE: CL2584 US NA
; CURRENT APPLICATION NUMBER: US/10/919,182
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: Patent version 3.2
; SEQ ID NO 8
; LENGTH: 369
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Mutant nitrilase B4
US-10-919-182-8

Alignment Scores:
Pred. No.: 1.42e-47 Length: 369
Score: 796.00 Matches: 160
Percent Similarity: 60.9% Conservative: 49
Best Local Similarity: 46.6% Mismatches: 130
Query Match: 42.3% Indels: 4
DB: 6 Gaps: 2

US-09-751-299-1 (1-1041) x US-10-919-182-8 (1-369)
```

QY 16 ACGAAGTATCGGCGCGCGGTCGAGAGCGCGCGGTCGTCGATCTCGACCGCA 75  
Db 6 SerlyspheleuAlaIaThrValGlnAlaGluProValITrpleuAspAlaAspAlaThr 25  
QY 76 GTCGAGAAAGCGATCGGCTGATCGAGAGCGCGCAAGCAAGCGTGGCTGATCGCA 135  
Db 26 IleAspYserIleGlyIleIleGlnIuAlaIaGlnYsglyAlaSerIleuIleAla 45  
QY 136 TTCGCAAGAGACTTGATTCGCGGCTATCCGCTTTTGATATAGCTGGGCGCGCTTGG 195  
Db 46 PheProGluValPheIleProGlyTyrrProTyrrPalatrrPleuGlyAspValIleCys 65  
QY 196 GGCATGCGCTTGGTCGACGCTATTTGAGAAATTCGTCGTCGCGGAGAGAGAGTGG 255  
Db 66 SerleuserPheThrSerArgTyrrHisGluAsnSerIleuGluLeuGlyAspArgMet 85  
QY 256 CAGGCGCTGGCGGATCGGCGCGCGCGCGCAAGCATGATTCGTGGCGCGCTATAGCGAG 315  
Db 86 ArgArgIleuGlnIleuAlaIaArgArgAsnIleuAlaIleuValMetGlyTyrSerGlu 105  
QY 316 CGCGCGGCGCGGACCTTATATAGGAGCGGCGATCTTGGCGCGCGATGGCATCTGATC 375  
Db 106 ArgGluAlaGlySerArgTyrrLeuSerGlnValPheIleAspGluArgGlyGluIleVal 125  
QY 376 GCGCGCGCGCGCAAGCTCAAGCTTACCCATCGGAGCGGACGATGTCGGAGGAGGAGAC 435  
Db 126 AlaAsnArgArgTyrrLeuTyrrProThrHisValGluArgThrIleTyrrGlyGluGlyAsn 145  
QY 436 GGCAGCATCTCGCGGTGACAGATACCGCATCGGCGCGCTCGGCGCGCTGTGTGTCG 495  
Db 146 GlyThrAspPheLeuThrHisAspPheAlaPheGlyArgValGlyGlyLeuAsnCyserTrp 165  
QY 496 GAGCAGATCCAGCATTTGTCGAATATCCCATGTATCGCGCGCGCAAGCAAGCTCAAGTC 555  
Db 166 GluHisPheGlnIleProLeuSerIleAlaIleuMetTyrrSerIleuGlyGluGlnValHisVal 185  
QY 556 GCGCGTGGCGGACCTTACGCTCATATCGGCGCATGCGCTATGGCTCGACCGAGAGTC 615  
Db 186 AlaSerTrpProAlaMetSerProLeuGlnProAspValPheGlnLeuSerIleGlnAla 205  
QY 616 AATACCGCGCAACGATCTACGCGGTGAGGCGGCGCTCATGTCGTCGTCGTCGTCGTC 675  
Db 206 AsnAlaThrValIleArgSerTyrrAlaIleGlnIleGlnThrPheValIleuCyserThr 225  
QY 676 GCGACGTTTGGCGGAGATATCAAGGTATTTGGTGTGATACCGCGCAAGAGATGTC 735  
Db 226 GlnValIleGlyProSerAlaIleGluThrPheCysLeuAsnAspGluGlnArgAlaIleu 245  
QY 736 CTCAAGCGCGCGCGGCTTTTGGCATATTTTCGCGCGCGCGCGCGCGCTTGGCGAG 795  
Db 246 LeuProGlnGlyCysGlyTrpAlaArgIleTyrrGlyProAspGlySerGluLeuAlaIle 265  
QY 796 CCGCTCCGCGAGACGGAAGAGGAGGAGCTGTCGTCGCGCATTCGATCGCATCGCG 855  
Db 266 ProleuAlaGluAspAlaGlnIleuTyrrAlaGlnIleAspLeuGlnIleu 285  
QY 856 TTGGCCAAAGCGCGCGGATTCGCGCGCGCATTTACGCGCGCAAGTAAAGCGCGCTG 915  
Db 286 LeuAlaIleValIleAlaAspProValGlyHisTyrrSerArgProAspValIleuSerVal 305  
QY 916 CTGCTGATCGACT-----CCGCGCAAGCGCTTCGATCGCTTGAAGCGCGATTC 966  
Db 306 GlnPheAspProAspAsnIleThrProValHisArg--IleGlyIleAspGlyArgLeu 324  
QY 967 GAACCGCAAAAGAGAGAGGAGGAGCGCGCGCGCTGCGGTGGAGGAGGAGCGC 1026  
Db 325 AspAlaIleThrArgSerArgValGluAsnPheArgLeuArgGlnAlaIleGlnIleGln 344  
QY 1027 GCGCGCGCGC 1035  
Db 345 ArgGlnAla 347  
RESULT 9

US-11-096-568A-20687  
; Sequence 20687, Application US/11096568A  
; Publication No. US20060048240A1  
; GENERAL INFORMATION:  
; APPLICANT: Alexandrov, Nikolai et al.  
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides  
; TITLE OF INVENTION: Theby  
; FILE REFERENCE: 2750-1592PUS2  
; CURRENT APPLICATION NUMBER: US/11/096,568A  
; CURRENT FILING DATE: 2005-04-01  
; NUMBER OF SEQ ID NOS: 34471  
; SEQ ID NO 20687  
; LENGTH: 333  
; TYPE: PRT  
; ORGANISM: Zea mays subsp. mays  
; FEATURE:  
; NAME/KEY: misc\_feature  
; LOCATION: (1)..(333)  
; OTHER INFORMATION: Cereb Seq. ID no. 12386686  
US-11-096-568A-20687  
  
Alignment Scores:  
Pred. No.: 2,23e-22 Length: 333  
Score: 430.50 Matches: 117  
Percent Similarity: 50.6% Conservative: 52  
Best Local Similarity: 35.0% Mismatches: 128  
Query Match: 22.9% Indels: 37  
Gaps: 8  
US-09-751-299-1 (1-1041) x US-11-096-568A-20687 (1-333)  
QY 16 ACGAAGTATCGGCGCGCGGTCGAGAGCGCGCGGTCGTCGATCTCGACCGCA 75  
Db 10 ThrThrAlaArgValIleThrValAlaGlnAlaSerSerValPheTyrrAspThrProAlaThr 29  
QY 76 GTCGAGAAAGCGATCGGCTGATCGAGAGCGCGCAAGCAAGCGTGGCTGATCGCA 135  
Db 30 LeuAspYserAlaGluYsglyLeuValAlaGlnAlaIleGlyTyrrGlySerGlnLeuValLeu 49  
QY 136 TTCGCAAGACTTGATTCGCGGCTATCC-----TTTGGATATGGCTGGGC 183  
Db 50 PheProGluValPheValGlyTyrrProHisGlySerThrPheGlyLeuValIleGly 69  
QY 184 ----GCGCGCTTGGCGGATGCGCTTGTGTCAGGCGCTATTTGAGAAATTCGCTGTCGC 240  
Db 70 AsnArgThrAlaYsglyYsglyAspPheGlnYsTyrrHisAlaSerAlaIleAspVal 89  
QY 241 GGCAGCAAGAGTGGCGAGCGCGCTGAGATGCGCGCGCGCGCATGATGTCGTG 300  
Db 90 ProGlyProGluValSerArgLeuSerAlaLeuAlaGlyTyrrYsValPheLeuVal 109  
QY 301 GCGCGCTATAGCGAGCGCGCGGCGGCGAGCTTATATGGCGCGAGATCTTGGCGCC 360  
Db 110 IleGlyValValGluArgAlaGlyTyrrThrLeuTyrrAsnThrValLeuSerPheAspPro 129  
QY 361 GATGGCATGTGATCGCGCGCGCGCGCGGAGCTCAAGCTTACCATCGGAGCGGACCGT 420  
Db 130 LeuGlyTyrrLeuGlyYsHisArgYsValMetProThrAlaLeuGluArgValPhe 149  
QY 421 TTGCGGAGGAGAGCGGCGGCGCATCTGCGGATGACATTCGCGCATCGGCGCGCTGGC 480  
Db 150 TrpGlyPheGlyAspGlySerThrIleProValTyrrAspThrProIleGlyYsMetGly 169  
QY 481 GCGCTGTGTGTCGGAGCAATTCAGCGCTTGTGCAATATACGCGATGATCGCGCGGAC 540  
Db 170 AlaLeuIleCyserTrpGluAsnArgMetProLeuLeuArgThrAlaMetTyrrAlaIleGly 189  
QY 541 GAACAGGTCCAGCTCGGTCGTGGCGGAGCTTACGCTTATGCGGAGATGGCTATGCG 600  
Db 190 IleGluIleTyrrCysAla-----TACGCGGTG 195  
QY 601 CTCGACCGGAGGTCAATACCGCGCAAGCCAGATC-----TACGCGGTG 645

```
Dh 196 -----ProthrValaAspCysMetProthrTripleuSerSerMetThrHisIleAlaLeu 213
Qy 646 GAGGGCGGCTGCTACGTCGTCGCGCTGCTGCGCG-----ACCGTT 684
Db 214 GluIdylGlyCysPheValLeuSerAlaCysGlnPheCysArgAlaGlyAsnTyrProPro 233
Qy 685 TCGCGGAGATGATCAAGGTA---TTGGTGGATACGCCCGCAAGAGATGTCCTCAAG 741
Db 234 ProProGluTyrThrPheCysGlyLeuGluGluGluProSerProGluSerValValCys 253
Qy 742 GCCGCGCGCGCTTTGCCATGATTTTGGCGCCGACGCGCGCGCTTGGCGGACCGCTC 801
Db 254 SerIdylGly-----SerValIleIleSerProLeuGlyThrValIleuAlaGlyProAsn 271
Qy 802 CCGGAGACCGGAAGGAGGATGCTGTGTCGGATATCGACCTCGGCGATGATGCGTTGGCC 861
Db 272 TyrGluSer---GluAlaLeuThrAlaAspLeuAspLeuGlyGluIleValArgAla 290
Qy 862 AAGCGGCGGCGCGATCCGCGCGGCGCACTATTTCAGCGCCGACGTACGCGGCTGCTGCTG 921
Db 291 LysPheAspPheAspValValGlyHisTyrSerArgProGluValIleuSerLeuValVal 310
Qy 922 GATGACGTCCTCGCGCCCAACGCTGTCACGCTTGATGCCGCA 963
Db 311 LysSerAspProLysProAlaValSerPheIleSerAlaIa 324

RESULT 10
US-11-096-568A-20686
; Sequence 20686, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nikolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 20686
; LENGTH: 351
; TYPE: PRT
; ORGANISM: Zea mays subsp. mays
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)-(351)
; OTHER INFORMATION: Ceres Seq. ID no. 12386685
US-11-096-568A-20686

Alignment Scores:
Pred. No.: 2,24e-22 Length: 351
Score: 430.50 Matches: 117
Percent Similarity: 50.6% Conservative: 52
Best Local Similarity: 35.0% Mismatches: 128
Query Match: 22.9% Indels: 37
DB: 7 Gaps: 8

US-09-751-299-1 (1-1041) x US-11-096-568A-20686 (1-351)
Qy 16 ACCGAATATCGCGCGCGCGGTGTCAGAGCGCGCGCTGTCCTCGATTCGACCGCACA 75
Db 28 ThrThrAlaArgValThrValValGlnAlaSerSerAlaPheTyrArgThrProAlaThr 47
Qy 76 GTGAGAAACCGATCGCGCTGATCGACAGCGCGCGCAAGACGCTGCGCGCTGATCGCA 135
Db 48 LeuAspLysAlaGluLysLeuValAlaGluAlaIleGlyTyrGlySerGlnLeuValLeu 67
Qy 136 TTCGCGAGACTTGATTCGCGGCTATCC-----TTTGGATATGGCTGGCG 183
Db 68 PheProGluValAlaPheValGlyGlyTyrPheHisGlySerThrPheGlyLeuValValGly 87
Qy 184 ---CGCGCGGCTTGGGCGATGCGCTTGCACGCGCTATTTCAGGAATTCGCTGCGCGC 240
Db 88 AsnArgThrAlaLysGlyLysGluAspPheGlnLysTyrHisAlaSerAlaIleAspVal 107
```

```
Qy 241 GCGACAGACAGTGGCAGAGCCCTTGGCGAGATGCGCCCGCGCGACGCGCATGATGTCG 300
Db 108 ProGluProGluValSerArgLeuSerAlaLeuAlaGlyLysTyrValPheLeuVal 127
Qy 301 GCCGCTATATCCAGAGCGCGCGCGCGAGCTCTATATGAGCGAGGATCTTCGCGCCC 360
Db 128 IleGlyValValGluArgAlaGlyTyrThrLeuTyrAsnThrValLeuSerPheAspPro 147
Qy 361 GATGCGATGTATCGCGCGCGCGCGCGAGCTCAAGCTTACCATCGACGCGACCGCTG 420
Db 148 LeuGlyLysTyrLeuGlyLysHisArgLysValMetProThrAlaLeuGluValPhe 167
Qy 421 TTCGCGAGGAGAGAGCGAGCCATCTCGCGGTGTCAGATTCGCGCATCGGCGCTCGC 480
Db 168 TrpGlyPheGlyAspGlySerThrIleProValTyrAspThrProIleGlyLysMetGly 187
Qy 481 GCGCTGTGTTGCTGGAGACATCCAGCCATTGTGCAAAATACGCATGTACGCGCGCAC 540
Db 188 AlaLeuIleCysTrpGluAsnArgMetProLeuLeuArgThrAlaMetTyrAlaLysGly 207
Qy 541 GAACAGGTCCAGCTCGCGCTGTCGCGCGAGCTTACGCTTATGCGCGCATGCGCTATGCG 600
Db 208 IleGluIleTyrCysAla----- 213
Qy 601 CTCGAGACCGGAGGTCAATACCGCGCGACGACAGATC-----TACGCGGTC 645
Db 214 -----ProthrValaAspCysMetProthrTripleuSerSerMetThrHisIleAlaLeu 231
Qy 646 GAGGCGGCTGCTACGTCGTCGCGCTGTCGCGCG-----ACCGTT 684
Db 232 GluGlyGlyCysPheValLeuSerAlaCysGlnPheCysArgArgLysAsnTyrProPro 251
Qy 685 TCGCGGAGATGATCAAGGTA---TTGGTGGATACGCCCGCAAGAGATGTTCTCAAG 741
Db 252 ProProGluTyrThrPheCysGlyLeuGluGluGluProSerProGluSerValValCys 271
Qy 742 GCCGCGCGGCTTTGCCATGATTTTGGGCGCGACGCGCGCGCTTGGCGGACCGCTC 801
Db 272 SerGlyGly-----SerValIleIleSerProLeuGlyThrValLeuAlaGlyProAsn 289
Qy 802 CCGGAGACCGGAAGGAGATGCTGTCGTCGCGCGATATGACACTCGGCAATGCGGTTGGCC 861
Db 290 TyrGluSer---GluAlaLeuThrAlaAspLeuAspLeuGlyIleValArgAla 308
Qy 862 AAGCGCGCGCGCATCCGCGCGGCGCACTATTTCAGCGCGCGACGTACGCGGCTGCTGCTG 921
Db 309 LysPheAspPheAspValValGlyHisTyrSerArgProGluValIleuSerLeuValVal 328
Qy 922 GATGACGTCCTCGCGCCCAACGCTGTCACGCTTGATGCCGCA 963
Db 329 LysSerAspProLysProAlaValSerPheIleSerAlaIa 342

RESULT 11
US-11-096-568A-6955
; Sequence 6955, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nikolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 6955
; LENGTH: 330
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)-(330)
; OTHER INFORMATION: Ceres Seq. ID no. 15169318
```



```

Db      191 CyETripLubAsnArgMetProLeuLeuArgThrAlaMetTyrAlaIysGlyValGluIle 210
QY      550 CAGGTGGCGTGGTGGCGGATTCAGCTCTATGCGGCGATGCGCTTATGCGTGGACCG 609
Db      211 TyrCyAla-----ProThr-----AlaAspAlaArg 219
QY      610 GAGGTCAATACCGCGGACGAGCATGTACGCGGTGCGAGGCGGCTGTACGTGGCG 669
Db      220 AspValTrrpGlnAlaSerMetThrIleIleAlaLeuGlnGlyGlyCysPheValLeuSer 239
QY      670 TCG-----TGGCGG-----ACGTTTCGCGGAGATGATC-----AAG 702
Db      240 AlaAsnGlnPheCysArgGArgArgAspTyrProProProGlnGlyTyrValPheAlaGly 259
QY      703 GTATTGGTGGAT---ACGCGCGACAGAGAGATGTCTCAAGCGCGGCGGCTTTTGGC 759
Db      260 ThrGlnValaAspLeuThrProAspSer-----ValValCysAlaGlyGly-----Ser 275
QY      760 ATGATTTTCGGGCGCGGACGCGCGCTTGGCGGACGCGCGCGGACGAGAGAGGA 819
Db      276 ValIleIleSerProLeuGlyAlaValIleuAlaGlyProAsnTyrAsp---GlyGluAla 294
QY      820 CTGTGTCGCGCGATATGCACTGCGCATGATCGCGTGGCGCAAGCGCGCGCGATCCG 879
Db      295 LeuIleSerAlaAspLeuAspLeuGlyGlnIleAlaArgAlaIysPheAspPheAspVal 314
QY      880 GCGGCGCCTATTTCACGCGCGCGGACGTAAAGCGCGCTGCTGCTGATGACGTCGCGGCCAA 939
Db      315 ValGlnIleIstYrSerArgProGlnValaLeuSerLeuThrValIysAspPheIleProThrAsn 334
QY      940 GCGGTCGTCACGCTTGATGCGCATTCGACCGCGCAAAAGGAGAGCAAG 987
Db      335 ProValIleThrPheThrSerThrSerThrIstIleGlnAspIysThrLys 350

RESULT 13
US-11-096-568A-6956
; Sequence 6956, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nikolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 6956
; LENGTH: 233
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(233)
; OTHER INFORMATION: Ceres Seq. ID no. 15169319
US-11-096-568A-6956

Alignment Scores:
Pred. No.: 1,99E-14 Length: 233
Score: 315.00 Matches: 88
Percent Similarity: 51.6% Conservative: 40
Best Local Similarity: 35.5% Mismatches: 94
Query Match: 16.7% Indels: 26
DB: 7 Gaps: 9

US-09-751-299-1 (1-1041) x US-11-096-568A-6956 (1-233)

QY      274 GCCCGCGCGACGCGCATGCTGTCGCGCGCTATGACGCGCGCGCGACGCTC 333
Db      2 AaaGlyLysTyrTyrValIstIleuValIleMetGlyValIleGlnIleArgAspIlyTrrhIleu 21
QY      334 TATATGGCGCGACGCGATCTTTCGCGCGCGGATGCGGATGTACGCGCGCGCGCGCAAGCTC 393
Db      22 TyrCysThrValIleuPheAspSerGlnGlnIstIleuGlyLysIstIleuGlyIstIle 41

```

```

QY      394 AACCTTACCATTGCGGACGCGACCGCTTTGCGCGAGGAGAGACGCGACCATTCGCGGTG 453
Db      42 MetProThrAlaLeuGlnIleArgValIleTrrpGlyPheGlyAspGlySerThrIleProVal 61
QY      454 CAGGATACCGCATCGCGCGCTCGCGCGCTGTGTTGCTGCGGACACATCCAGCATTG 513
Db      62 PheGlnThrProValGlyLysIleGlyAlaIleValIleCysTrrpGlnIleAsnArgMetProLeu 81
QY      514 TCGAAATACCGCATGTAGCGCGCGACGAGACAGTTCACGTCGCTCGTGGCGCGCTTC 573
Db      82 LeuArgThrAlaMetTyrAlaIysGlyValGlnIleTyrCysAla-----ProThr--- 98
QY      574 AGCTCTATCGCGGATGCGCTATGCGCTGCGGACCGGAGTCAATACCGCGCGCAAGCG 633
Db      99 -----AlaAspAlaAspValTrrpGlnAlaSerMetThr 110
QY      634 ATCTACGCGGTGAGGCGCGCTGTACTGCTGGCGTGC-----TGGCGG----- 678
Db      111 HisIleAlaLeuGlnGlyGlyCysPheValIleuSerAlaAsnGlnPheCysArgArgArg 130
QY      679 -----ACGTTTCGCGGAGATGATC-----AAGGTATGTGGAT---ACGCGCGAC 723
Db      131 AspTyrProProProGlnGlyTyrValPheAlaGlyThrGlnValaAspLeuThrProAsp 150
QY      724 AAGGAGATGTCTCAAGCGCGCGGCTTTTGCATGATTTTGGCGCGCGCGCGC 783
Db      151 Ser-----ValValCysAlaGlyGly-----SerValIleIleSerProLeuGlyAla 166
QY      784 GCGCTGCGCGGCGCGCTCGCGGACGAGAGGAGCTGTGATGCGCGCATTCGACCTC 843
Db      167 ValIleuAlaGlyProAsnTyrAsp---GlyGluAlaLeuIleSerAlaAspLeuAspLeu 185
QY      844 GCGATATGCGCTTGGCGCAAGCGCGCGCATTCGCGCGCGCATTTATTCAGGCGCGAC 903
Db      186 GlyGlnIleAlaArgAlaIysPheAspPheAspValIleGlyIstIleArgProGln 205
QY      904 GTAACGCGCTGCTGCTGATGACGTCGCGCGCAAGCGGTGTACGCTTATGCGCGCA 963
Db      206 ValLeuSerLeuThrValIysAspPheIstProThrAsnProValThrPheThrSerThr 225
QY      964 TTCGAACCGCAAAAGGAGAGCAAG 987
Db      226 ThrIstIleGlnAspIysThrLys 233

RESULT 14
US-11-096-568A-20688
; Sequence 20688, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nikolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 20688
; LENGTH: 193
; TYPE: PRT
; ORGANISM: Zea mays subsp. mays
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)..(193)
; OTHER INFORMATION: Ceres Seq. ID no. 12386687
US-11-096-568A-20688

Alignment Scores:
Pred. No.: 2.7E-12 Length: 193
Score: 284.00 Matches: 75
Percent Similarity: 51.0% Conservative: 28
Best Local Similarity: 37.1% Mismatches: 67
Query Match: 15.1% Indels: 32

```



```

Db      353  gGlyAlaSerAlaValAlaAla-----SerArgProThrProSerSerLalAl 379
QY      161  TAGCCGGGAATCCAAAGTCTCTGGGAATGCGATCAGCGCGACAGCTGCTGGCGCGCTGC 102
        |||  ::|  |||  |||||  |||  |||
Db      379  aserThrGlyThrHisProAlaGly-----GlyAlaProAlaValHisProGlyAl 396
QY      101  TCGATCAGGCGCGATCGCTTTCGACTGTGCGGTGCGAGTCCAGATCCAGAGAACACCGCGGGGC 42
        |||  ::|  |||||  |||||  |||||  |||
Db      386  aAlaAlaGlyArg-----GlyArgGlyArgGlyAlaArgAlaArgAl 410
QY      41  TGCACCGCGCGCGC 29
        |||||
Db      410  uGlyArgProArg 414

```

Search completed: April 27, 2006, 01:53:07  
Job time : 29.2628 secs



GenCore version 5.1.7  
(c) 1993 - 2006 Biocore

9934.641 Million cell updates/sec

Sequence: 1 atgtcggagcccatgaagaa.....gcgccgcgcgcgcagtag 1041

Gapop 10.0 , Gapext 1.0

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

### Listing first 45 summaries

Database : EST: \*

```

1:  gb_est1:*
2:  gb_est2:*
3:  gb_est3:*
4:  gb_htc:*
5:  gb_est4:*
6:  gb_est5:*
7:  gb_est6:*
8:  gb_est7:*
9:  gb_gss1:*
10: gb_gss2:*
11: gb_gss3:*

```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
C 1	155	14.9	888	9	AQ329770	AQ329770 nbx00045B
2	153	14.7	622	9	AQ329769	AQ329769 nbx00045B
3	140.8	13.5	673	8	DR613467	DR613467 EST102159
4	140.8	13.5	732	8	DR613467	DR613467 EST102139
5	140.8	13.5	810	8	DR629486	DR629486 EST101961
6	140.8	13.5	841	8	DR612130	DR612130 EST102225
7	140.8	13.5	856	8	DR622046	DR622046 EST101217
8	140.8	13.5	858	8	DR610522	DR610522 EST102065
9	129.8	12.5	655	6	CF453420	CF453420 ff536_1_I
10	129.4	12.4	840	7	CN808140	CN808140 B100D_EST
11	129.4	12.4	998	7	CN808242	CN808242 B100D_EST
12	127.6	11.3	769	8	DR609823	DR609823 EST99951
13	122.6	11.2	840	8	DR709497	DR709497 Aen_11063
14	116.6	11.8	580	9	AQ160684	AQ160684 mgx000061
15	109.4	10.5	804	8	DR709485	DR709485 Aen_11049
C 16	108	10.4	772	6	CD459513	CD459513 P509_03j1
17	101.2	9.7	551	7	CO149108	CO149108 EST824161
18	101.2	9.7	553	7	CO149186	CO149186 EST824239
19	99.6	9.6	610	7	CO134433	CO134433 EST829104
20	97.8	9.4	546	7	CO139118	CO139118 EST833789
21	91	8.7	536	8	DR622501	DR622501 EST101262
22	91	8.7	561	8	DR622709	DR622709 EST101283

23	91	8.7	595	8	DR626172	EST101630
24	91	8.7	633	8	DR622498	EST101262
25	91	8.7	653	8	DR626120	EST101624
26	91	8.7	758	8	DR625821	EST101594
27	90.6	8.7	515	7	CO137063	CO137062 EST831734
28	88.2	8.6	409	1	AW711275	fai1ine.f
29	88.6	8.5	866	8	DR621807	DR621807 EST101193
30	88.2	8.5	434	1	AF408416	AF408416 AF408416
31	83.8	8.0	937	8	DR624629	DR624629 EST101475
32	83.6	8.0	903	8	DR630423	DR630423 EST102055
33	81	7.8	532	2	BG278308	BG278308 adoznp.r
34	80.8	7.8	838	10	C8381199	C8381199 FSAN59TR
35	80.2	7.7	418	1	AW713376	AW713376 g8C04ne.f
36	75.4	7.2	417	8	AW710819	AW710819 e6h01ne.f
37	75.4	7.2	672	1	DR629554	DR629554 EST101968
38	74.2	7.1	441	10	CM493581	CM493581 fdb0001f2
39	74.2	7.1	546	10	CM149572	CM149572 104_548.1
40	72.6	7.0	383	8	DR625648	DR625648 EST101577
41	72.4	7.0	419	2	BF072824	BF072824 NCSPTC3t3
42	71.4	6.9	412	1	AI1329382	AI1329382 DsAlone.f
43	70.6	6.8	359	1	AW710710	AW710710 e5f03ne.f
44	68.8	6.6	744	7	CN128980	CN128980 RH0H1_32
45	68.8	6.6	747	7	CN124733	CN124733 RH0H1_6_C

## ALIGNMENTS

RESULT 1	LOCUS	DEFINITION
AQ329770/c	AQ329770	888 bp DNA linear
	nbxb0045f09r	CUCI Rice BAC library Oryza sativa (japonica
	cultivar-group)	genomic clone nbxb0045f09r, genomic survey

ACCESSION	AQ329770
VERSION	AQ329770.1
KEYWORDS	GI:4121620
SOURCE	GSS.
	<i>Oryza sativa</i> (japonica cultivar-group)

SOURCE ORGANISM	PHYTOCHEMICAL
<i>Oryza sativa</i> (japonica cultivar-group)	Phenol
<i>Oryza sativa</i> (japonica cultivar-group)	Phenol
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta	Phenol
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzeae; <i>Oryza</i> .	Phenol

REFERENCE	1 (bases 1 to 888)
AUTHORS	Wing, R.A. and Dean, R.A.
TITLE	A BAC End Sequencing Framework to Sequence the Rice Genome
JOURNAL	Unpublished (1998)
COMMENT	Contact: Wing RA

```

Email: twining@lemonsu.edu
Seg primer: GGAACACTATGACCATG
Class: BAC ends
High quality sequence start: 13
High quality sequence stop: 225
Location/Qualifiers
1: .888
FEATURES
source
```

## FEATURES

organisms="Oryza sativa (japonica cultivar-group)"  
/mol\_type="genomic DNA"  
/cultivar="japonica"  
/cultivar="Nipponbare"  
/db\_xref="taxon:39947"  
/clone="nxb00045f09r"  
/tissue\_type="leaf"  
/lab\_host="E. coli DH10B"  
/clone\_lib="CGI Rice BAC Library"  
/note="Vector: pBel0BAC11; Site 1: HindIII; Site 2:  
HindIII; Rice is one of two most popular grains in the  
world. Half of the world population especially those  
inhabiting highly populated areas of the humid tropics  
and subtropics, rely on rice as their primary source of

carbohydrate. Monocotyledonous rice is a diploid plant (2n=24) with a haploid genome equivalent of 431 Mbp (Arumuganathan and Earle, 1991). The relatively small genome of rice, three times larger than that of Arabidopsis, makes it suitable for genomic studies. In order to facilitate positional cloning, physical mapping and genome sequencing of rice, we have constructed a BAC library from *Oryza sativa*, Nipponbare variety. The library contains 36,864 clones with an average insert size of 128.5 kb providing 10.9 haploid genome equivalents. The deep coverage allows the isolation a particular sequence with a probability of 99.9 %. Two high density filters, each containing 18,432 clones (doubly spotted), represent the whole library for colony screening."

## ORIGIN

Query Match 14.9%; Score 155; DB 9; Length 888;  
Best Local Similarity 60.6%; Pred. No. 2.6e-25;  
Matches 251; Conservative 0; Mismatches 163; Indels 0; Gaps 0;

150 GATTCGGGCTATCCCTTTGGATATGCTGGGCGCGGCTTGGGGCATGGCTTCGT 209  
Db 414 GCTAACGGGATATTCATATTTAGCATAGTAGAGATTCTAGAAATGGGGAAGCATTA 355  
Qy 210 CCAGCGCTATTTCGAGATTGCTGTCGGCGGAGCAAGCAGTGCAGGCTTGGCGGA 269  
Db 354 CGATCGACAGTCAGAAAAACAGCANGNCCGAGATGCGCCGGAAGTCGGMATCGCGGA 295  
Qy 270 TGGGGCGCGCGCGGCGCATGTCATGTCGCGCGCTATATAGCGCGCGCGCGGAG 329  
Db 294 GCGCGCGCTGCGGACGAGCATCTGATCAGTCTGGGCTATAGCGAGCGCGCGGAG 235  
Qy 330 CCTTATATGGGCGGAGCGATCTTGGCGCGCGGATCGATTCGCGCGCGCGCAA 389  
Db 234 CCTTATATGCTGCTCACTGATGATGATGATGATGATGATGATGATGATGATGAT 175  
Qy 390 GCTTACGCTTACCATTCGAGCGGACCGCTGTTGGCGAGGAGACGGACCATCTGCG 449  
Db 174 GCTTACGCTTACCATTCGAGCGGACCGCTGTTGGCGAGGAGACGGACCATCTGCG 115  
Qy 450 GGTGCAAGTATCGGCGCATCGGCGCGCTGTTGGCGAGGAGACCATTCAGCG 509  
Db 114 GGTGCTGCTGAGGACCTTGGCGCGCTGTTGGCGAGGAGACCATTCAGCG 55  
Qy 510 ATTGTCGAAATGCGCATGATGACCGCGGAGCAAGCATGCTGCTGCTG 563  
Db 54 GCTTACCAATATGCTGATGATGATGATGATGATGATGATGATGATGATGATGAT 1

RESULT 2  
AQ329769 622 bp DNA linear GSS 08-JAN-1999  
LOCUS nbxb0045P09f CUGI Rice BAC Library *Oryza sativa* (japonica  
DEFINITION cultivar-group) genomic clone nbxb0045P09f, genomic survey  
sequence.  
VERSION AQ329769  
KEYWORDS GSS.  
SOURCE *Oryza sativa* (japonica cultivar-group)  
ORGANISM *Oryza sativa* (japonica cultivar-group)  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Ehrhartoideae; Oryzaceae; *Oryza*.  
1 (bases 1 to 622)  
Wing, R.A. and Dean, R.A.  
A BAC End Sequencing Framework to Sequence the Rice Genome  
Unpublished (1998)  
Contact: Wing RA  
Clemson University Genomics Institute  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Tel: 864 656 7288  
Fax: 864 656 4293

Email: rwing@clemson.edu  
Seq primer: TAATACGACTCACTATAGG  
Class: BAC ends  
High quality sequence stop: 423.  
Location/Qualifiers

## FEATURES

1..622  
/organism="Oryza sativa (japonica cultivar-group)"  
/mol\_type="genomic DNA"  
/cultivar="japonica"  
/cultivar="Nipponbare"  
/db\_xref="taxon:39947"  
/clone="nbxb0045P09f"  
/tissue\_type="leaf"  
/lab\_host="E. coli DH10B"  
/clone\_lib="CUGI Rice BAC Library"  
/note="Vector: pBeloBAC11; Site 1: HindIII; Site 2:  
HindIII; Rice is one of two most popular grains in the  
world. Half of the world population especially those  
inhabiting highly populated areas of the humid tropics  
and subtropics, rely on rice as their primary source of  
carbohydrate. Monocotyledonous rice is a diploid plant  
(2n=24) with a haploid genome equivalent of 431 Mbp  
(Arumuganathan and Earle, 1991). The relatively small  
genome of rice, three times larger than that of  
Arabidopsis, makes it suitable for genomic studies. In  
order to facilitate positional cloning, physical mapping  
and genome sequencing of rice, we have constructed a BAC  
library from *Oryza sativa*, Nipponbare variety. The  
library contains 36,864 clones with an average insert size  
of 128.5 kb providing 10.9 haploid genome equivalents. The  
deep coverage allows the isolation a particular sequence  
with a probability of 99.9 %. Two high density filters,  
each containing 18,432 clones (doubly spotted), represent  
the whole library for colony screening."

## ORIGIN

Query Match 14.7%; Score 153; DB 9; Length 622;  
Best Local Similarity 60.9%; Pred. No. 7.5e-25;  
Matches 249; Conservative 0; Mismatches 160; Indels 0; Gaps 0;

Qy 577 CTCTATCGGGGCGATGCGCTTATGCGCTCGGACGAGAGTCAATACCGCGCAAGCATG 636  
Db 5 CTCTACAAAGGAGAAAGGCTTTCGCGCTGCTCCGAGCTCAACAAATGCTTCAGCGAGT 64  
Qy 637 TAGCGGCTGAGAGGCGGCTGCTACGTCGTCGCGTGGCGGACCGTTTCGCGGAGATG 696  
Db 65 TAGCGGCTGAGAGGCGGCTGCTACGTCGTCGCGTGGCGGACCGTTTCGCGGAGATG 124  
Qy 697 ATCAAGGTATGATGATACGCCCAAGAGATGTTCTCAAGGCGCGCGGTTT 756  
Db 125 ATCAAGGTATGATGATACGCCCAAGAGATGTTCTCAAGGCGCGCGGTTT 184  
Qy 757 GCATGATTTTGGGCGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 816  
Db 185 GCATGATTTTGGGCGGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 244  
Qy 817 GGAATGCTGTCGCGGATATCGACCTCGGATGATGATGATGATGATGATGATGATGAT 876  
Db 245 GCGCTGATGTCGGAATGATGATGATGATGATGATGATGATGATGATGATGATGAT 304  
Qy 877 CCGGCGGCGCATATTCACGCGCGGACGTAACGCGGCTGCTGCTGATGATGATGATG 936  
Db 305 CCGGCGGCGCATATTCACGCGCGGACGTAACGCGGCTGCTGCTGATGATGATGATG 364  
Qy 937 CAAGCGTGTGACGCTTATGATGATGATGATGATGATGATGATGATGATGATGATG 985  
Db 365 GCGCGGCTGAGCGGCTTATGATGATGATGATGATGATGATGATGATGATGATGATG 413

RESULT 3  
DR631467 699 bp mRNA linear EST 11-JUL-2005  
LOCUS DR631467  
DEFINITION EST1021595 Fv1 Giberella moniliformis cDNA clone FvIB956, mRNA

sequence.  
 DR631467  
 VERSION  
 DR631467.1 GI:70706307  
 EST.  
 SOURCE  
 Gibberella moniliformis  
 ORGANISM  
 Gibberella moniliformis  
 Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
 Hypocreomycetidae; Hypocreales; Nectriaceae; Gibberella.  
 REFERENCE  
 1 (bases 1 to 699)  
 Brown, D.W., Cheung, F., Proctor, R.H., Butchko, A.E., Zheng, L., Lee, Y.,  
 Utebäck, T., Smith, S., Feldblyum, T., Glenn, A.E., Plattner, R.D.,  
 Kendra, D.F., Town, C.D. and Whitelaw, C.A.  
 Analysis of 87,000 expressed sequence tags reveals alternatively  
 spliced introns in multiple genes of the fumonisin gene cluster  
 Unpublished (2005)  
 JOURNAL  
 Contact: Brown, D.W.  
 COMMENT  
 USDA/ARS/NCAUR  
 1815 N. University St, Peoria, IL 61604, USA  
 Tel: 309 681 6230  
 Fax: 309 681 6689  
 Email: brown@ncaur.usda.gov  
 TIGR sequence name: FVIE956TH  
 Seq primer: AAT TAA CCC TCA AAG GG.  
 Location/Qualifiers  
 1..699  
 /organism="Gibberella moniliformis"  
 /mol\_type="mRNA"  
 /strain="m3125"  
 /db\_xref="taxon:117187"  
 /clone="FVIE956"  
 /cissue\_type="mycelia"  
 /clone\_lib="FVI"  
 /note="Vector: pBluescript II SK(+) XR; Site 1: EcoRI;  
 Site 2: XhoI; anamorph: Fusarium verticillioides. Library  
 FVI was prepared from growth on excised maize seedling  
 roots and shoots. The roots and shoots were prepared as  
 described above and then inoculated by dipping briefly in  
 a suspension of 5 x 10<sup>6</sup> conidia per ml. The inoculated  
 roots and shoots were then incubated on moistened Whatman  
 #1 filter paper at room temperature under sterile  
 conditions for 4 days. Fungal mycelial was collected,  
 frozen in liquid nitrogen, ground to a powder, and then  
 added to TRIzol Reagent (Invitrogen, Carlsbad CA) at  
 approximately 1 g mycelia per 10 ml TRIzol. The cDNA was  
 directionally ligated into the pBluescript II SK(+) XR  
 vector (CDNA Synthesis Kit; Stratagene)."

## ORIGIN

Query Match 13.5%; Score 140.8; DB 8; Length 699;  
 Best Local Similarity 54.2%; Pred. No. 5e-22;  
 Matches 308; Conservative 0; Mismatches 257; Indels 3; Gaps 1;

QY 9 GCCCATGACGAAGATCGCGCGCGGCGGAGCGCGCGGCTTCTTCATCGA 68  
 Db 53 GGCATATACCAAGTCAAGAGCGCGCTGCTGCTCACTCCGAGCTTGATGATCTTGA 112  
 QY 69 CCGCACAGTGAAGAACGATCGGCTGATCGAGCAGCGGCCAAGCAGACGTCGCCCT 128  
 Db 113 GGGGGGTTTCGCAAGCATGCACTTCATCAACAGAGCGCGGCAAGCTGGTCGAAACT 172  
 QY 129 GATCGCATTCGCGAGAGCTTGATTCGCGGCTATCCCTTTTGATATGCGCGCGGCC 188  
 Db 173 CGTCGCTTTCCGAAAGTTTGATTCCTGATATCTTACTGATGAGGAAAGTCACTTA 232  
 QY 189 GCGTTGGGCGATGCGCTTCGTCAGCGCTATTTGAGAATTCGCTCGCGCGCGCA 248  
 Db 233 CTCCTCATTCCTCCCATGCTGAAGCGCTACCGGAGAACTCCATGCGCGCTCACTTGA 292  
 QY 249 GCAATGCGAGGCGCTGCGGAGTGGCGCGCGCGCGCAGGAGTGCATGCTGCGCGCTA 308  
 Db 293 GGAATATCGCGCTATTCGTCGCGAGCGCGGATATCAAGATCTACGTCCTCCGCGCTT 352

QY 309 TAGCGAGCGCGCGGCGGCGAGCCTCTATATAGGCGAGCGCATCTTCCGCCGATGCGGA 368  
 Db 353 CTCGAGATTTGACACGCAACTCTTTACCTCGGCCAAGTCTTATGGCCGCGATGGCTC 412  
 QY 369 TCTGATGCGCGCGCGCGGCAAGCTCAAGCTTACCATGCGAGCGCACCGTTCGCGGA 428  
 Db 413 GGTCAATCAACCAACCGCTGCGCAAGATCAAGCACTCAAGTTGAGAAAGCTTTATACGCGGA 472  
 QY 429 GGGAGACGG---GAGCATTCGCGGTGACAGATACCGCATGGGCGGCGCGCT 485  
 Db 473 TGCTCCGGGAGATCTTATGAGCCGCTTACGAGACTGACATCGGCGGTGTGCGAGCT 532  
 QY 486 CTGTGCTGGAGACATCCAGCATTTGCAATAGCCCATGATGAGCGCGCGGACCAACA 545  
 Db 533 TAATGCTGGAGAAATGAAACCCCTTCCCAAGTCTCTCAAGTTTCTGCTGTAGCA 592  
 QY 546 GGTCAAGTTCGCTGTCGCGAGCTTC 573  
 Db 593 GGTTCACATGCTGCTGGCGCGCTTAC 620

RESULT 4  
 DR631867  
 LOCUS  
 EST1021995 FVI Gibberella moniliformis cDNA clone FVIE80, mRNA  
 DEFINITION  
 sequence.  
 DR631867 732 bp mRNA linear EST 11-JUL-2005  
 DR631867.1 GI:70706707  
 VERSION  
 DR631867  
 EST.  
 SOURCE  
 Gibberella moniliformis  
 ORGANISM  
 Gibberella moniliformis  
 Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
 Hypocreomycetidae; Hypocreales; Nectriaceae; Gibberella.

REFERENCE  
 1 (bases 1 to 732)  
 Brown, D.W., Cheung, F., Proctor, R.H., Butchko, A.E., Zheng, L., Lee, Y.,  
 Utebäck, T., Smith, S., Feldblyum, T., Glenn, A.E., Plattner, R.D.,  
 Kendra, D.F., Town, C.D. and Whitelaw, C.A.  
 Analysis of 87,000 expressed sequence tags reveals alternatively  
 spliced introns in multiple genes of the fumonisin gene cluster  
 Unpublished (2005)  
 JOURNAL  
 Contact: Brown, D.W.  
 COMMENT  
 USDA/ARS/NCAUR  
 1815 N. University St, Peoria, IL 61604, USA  
 Tel: 309 681 6230  
 Fax: 309 681 6689  
 Email: brown@ncaur.usda.gov  
 TIGR sequence name: FVIE80TH  
 Seq primer: AAT TAA CCC TCA AAG GG.  
 Location/Qualifiers

## FEATURES

source

1..732  
 /organism="Gibberella moniliformis"  
 /mol\_type="mRNA"  
 /strain="m3125"  
 /db\_xref="taxon:117187"  
 /clone="FVIE80"  
 /cissue\_type="mycelia"  
 /clone\_lib="FVI"  
 /note="Vector: pBluescript II SK(+) XR; Site 1: EcoRI;  
 Site 2: XhoI; anamorph: Fusarium verticillioides. Library  
 FVI was prepared from growth on excised maize seedling  
 roots and shoots. The roots and shoots were prepared as  
 described above and then inoculated by dipping briefly in  
 a suspension of 5 x 10<sup>6</sup> conidia per ml. The inoculated  
 roots and shoots were then incubated on moistened Whatman  
 #1 filter paper at room temperature under sterile  
 conditions for 4 days. Fungal mycelial was collected,  
 frozen in liquid nitrogen, ground to a powder, and then  
 added to TRIzol Reagent (Invitrogen, Carlsbad CA) at  
 approximately 1 g mycelia per 10 ml TRIzol. The cDNA was  
 directionally ligated into the pBluescript II SK(+) XR  
 vector (CDNA Synthesis Kit; Stratagene)."

## ORIGIN



LOCUS DR62130 841 bp mRNA linear EST 11-JUL-2005  
 DEFINITION EST1022258 Fv1 Gibberella moniliformis cDNA clone FVIE179, mRNA  
 ACCESSION DR62130  
 VERSION DR62130.1 GI:70706970  
 KEYWORDS EST.  
 SOURCE Gibberella moniliformis  
 ORGANISM Gibberella moniliformis  
 Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
 Hypocreomycetidae; Hypocreales; Nectriaceae; Gibberella.  
 REFERENCE 1 (bases 1 to 841)  
 AUTHORS Brown, D.W., Cheung, F., Proctor, R.H., Butchko, A.E., Zheng, L., Lee, Y.,  
 Uterback, T., Smith, S., Feldblyum, T., Glenn, A.E., Plattner, R.D.,  
 Kendra, D.F., Town, C.D. and Whitelaw, C.A.  
 Analysis of 87,000 expressed sequence tags reveals alternatively  
 spliced introns in multiple genes of the fumonisin gene cluster  
 Unpublished (2005)  
 CONTACT: Brown, D.W.  
 JOURNAL USDA/ARS/NCAR  
 COMMENT 1815 N. University St, Peoria, IL 61604, USA  
 Tel: 309 681 6230  
 Fax: 309 681 6689  
 Email: brown@ncar.usda.gov  
 TIGR sequence name: FVIE179TH  
 Seq primer: AAT TAA CCC TCA AAG GG.  
 Location/Qualifiers  
 1..841  
 /organism="Gibberella moniliformis"  
 /mol\_type="mRNA"  
 /strain="m3125"  
 /db\_xref="taxon:117187"  
 /clone="FVIE179"  
 /cissue\_type="mycelia"  
 /clone\_id="Fv1"  
 /note="Vector: pBluescript II SK(+); XR; Site 1: EcoRI;  
 Site 2: XhoI; anamorph: Fusarium verticillioides. Library  
 Fv1 was prepared from growth on excised maize seedling  
 roots and shoots. The roots and shoots were prepared as  
 described above and then inoculated by dipping briefly in  
 a suspension of 5 x 10<sup>6</sup> conidia per ml. The inoculated  
 roots and shoots were then incubated on moistened Whatman  
 #1 filter paper at room temperature under sterile  
 conditions for 4 days. Fungal mycelial was collected,  
 frozen in liquid nitrogen, ground to a powder, and then  
 added to Trizol Reagent (Invitrogen, Carlsbad CA) at  
 approximately 1 g mycelia per 10 ml Trizol. The cDNA was  
 directionally ligated into the pBluescript II SK(+) XR  
 vector (cDNA Synthesis Kit; Stratagene)."

ORIGIN  
 Query Match 13.5%; Score 140.8; DB 8; Length 841;  
 Best Local Similarity 54.2%; Pred. No. 5e-22;  
 Matches 308; Conservative 0; Mismatches 257; Indels 3; Gaps 1;

QY 9 GCCCATGACGAATATGCGCGCGCGGTCAGAGCGCGGCTTCTCGATCTCGA 68  
 DB 32 GGCATATACCAAGTACAGAGCGCGCTGCTACCTCCGAGCTGATGATCTTGA 91  
 QY 69 CCGACAGTGAAGAAAGGATCGGCTGATGAGACGAGCGGCAAGCAGACGTGGCCT 128  
 DB 92 GGGGCGTGTGGCAAGACATGACTTCAATACAGAGCGCGCAAGCTGGCTCAAACT 151  
 QY 129 GATGCAATCCACAGACTTGATTCGGGCTATCCCTTTGATATGCTGGCGCGCC 188  
 DB 152 CGTGGCTTCCGAGATTGGATCCCGATATCTTACTGATGAGAAAGTCACTTA 211  
 QY 189 GCGTTGGGAGATGCGCTTCTCCAGCGCTATTTGAGATTGCTCGTGGCGGACGAA 248  
 DB 212 CCTCAATCTCTCCCATGCTGAAGCGCTACCGGAGAACTCCATGCGCGTCACTGA 271  
 QY 249 GCATGGCAGGCGCTGGCGGATGCGCGCGCGCAGCGGATGATGCTGCGCGGCTA 308

DB 272 GGAATGCGCGGCTATTCGTGCGCAGCCGCGATTAACAGATCTAGCTCTCCGCGCTT 331  
 QY 309 TAGCGAGCGCGCGCGCGGAGCCTCTATATGGCCAGGCAATCTTGGCCCGCATGGCGA 368  
 DB 332 CTCTGAGATTGACACGCAACTCTTTAACCCTGCGCAAGTCTATCGCCGATGGCTC 391  
 QY 369 TCTGATGCGCGCGCGCGCGCAAGCTCAAGCTTACCCATGCGGAGCGCACCGTTCGCGCA 428  
 DB 392 GGTCAATCAACACCGCTGCGCAAGATCAAGCCATCTCACTGTTGAGAGCTTTATACGGCA 451  
 QY 429 GGGAGACG---CAGGCATCTCGCGGTGACAGATACCGCCATCGGCGCGCTGCGCGCT 485  
 DB 452 TGGCTCGGGGATATCTTCAATGAGCCGTTAGCGAGACTGACATCGGCGGTGCGCGCT 511  
 QY 486 CTGTGCTGGGAGACATCTCAGCATTTGTGAAATACGATGATGAGCGCGCGACGACA 545  
 DB 512 TAACTGTGGGAAACATGAACCCCTTCCTCAAGTCTCAAGCTTGTGCTGTGAGCA 571  
 QY 546 GGTCCAGTGGCGGTGCGCGGAGCTTC 573  
 DB 572 GGTTCACATCTGCTTGGCGCGCTTAC 599

RESULT 7  
 LOCUS DR622046 856 bp mRNA linear EST 11-JUL-2005  
 DEFINITION EST1012174 Fv1 Gibberella moniliformis cDNA clone FVIA704, mRNA  
 sequence.  
 ACCESSION DR622046  
 VERSION DR622046.1 GI:70696694  
 KEYWORDS EST.  
 SOURCE Gibberella moniliformis  
 ORGANISM Gibberella moniliformis  
 Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
 Hypocreomycetidae; Hypocreales; Nectriaceae; Gibberella.  
 REFERENCE 1 (bases 1 to 856)  
 AUTHORS Brown, D.W., Cheung, F., Proctor, R.H., Butchko, A.E., Zheng, L., Lee, Y.,  
 Uterback, T., Smith, S., Feldblyum, T., Glenn, A.E., Plattner, R.D.,  
 Kendra, D.F., Town, C.D. and Whitelaw, C.A.  
 Analysis of 87,000 expressed sequence tags reveals alternatively  
 spliced introns in multiple genes of the fumonisin gene cluster  
 Unpublished (2005)  
 CONTACT: Brown, D.W.  
 JOURNAL USDA/ARS/NCAR  
 COMMENT 1815 N. University St, Peoria, IL 61604, USA  
 Tel: 309 681 6230  
 Fax: 309 681 6689  
 Email: brown@ncar.usda.gov  
 TIGR sequence name: FVIA704TH  
 Seq primer: AAT TAA CCC TCA AAG GG.  
 Location/Qualifiers  
 1..856  
 /organism="Gibberella moniliformis"  
 /mol\_type="mRNA"  
 /strain="m3125"  
 /db\_xref="taxon:117187"  
 /clone="FVIA704"  
 /cissue\_type="mycelia"  
 /clone\_id="Fv1"  
 /note="Vector: pBluescript II SK(+); XR; Site 1: EcoRI;  
 Site 2: XhoI; anamorph: Fusarium verticillioides. Library  
 Fv1 was prepared from growth on excised maize seedling  
 roots and shoots. The roots and shoots were prepared as  
 described above and then inoculated by dipping briefly in  
 a suspension of 5 x 10<sup>6</sup> conidia per ml. The inoculated  
 roots and shoots were then incubated on moistened Whatman  
 #1 filter paper at room temperature under sterile  
 conditions for 4 days. Fungal mycelial was collected,  
 frozen in liquid nitrogen, ground to a powder, and then  
 added to Trizol Reagent (Invitrogen, Carlsbad CA) at  
 approximately 1 g mycelia per 10 ml Trizol. The cDNA was  
 directionally ligated into the pBluescript II SK(+) XR

FEATURES  
 source



```

RESULT 9
LOCUS      CF453420          655 bp      mRNA      linear      EST 01-JAN-2004
DEFINITION fts36.1_124 Fusarium verticillioides fcc1 mutant subtraction
            library Gibberella moniliformis cDNA, mRNA sequence.
ACCESSION  CF453420
VERSION     CF453420.1  GI:40546865
KEYWORDS
SOURCE      EST.
            Gibberella moniliformis
            Gibberella moniliformis
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
            Hypocreomycetidae; Hypocreales; Nectriaceae; Gibberella.
REFERENCE   1 (bases 1 to 655)
AUTHORS     Pittella,A.M., McIntyre,L.M., Payne,G.A. and Moloshuk,C.P.
TITLE       Comparison of gene expression in the wild type and the fcc1 mutant
            of Fusarium verticillioides reveals expressed sequence tags
            associated with fumonisin biosynthesis
            unpublished (2003)
JOURNAL     Contact: Moloshuk CP
COMMENT     Department of Botany and Plant Pathology
            Purdue University
            915M, State Street, West Lafayette IN 47907-2054, USA
            Tel: 765 494 3450
            Fax: 765 494 0363
            Email: wloshuk@purdue.edu
            Insert Length: 655      Std Error: 0.00
            Plate: f6 row: 0 column: 11
            Seq primer: 77.
FEATURES
source      1..655
            location/Qualifiers
            /organism="Gibberella moniliformis"
            /mol_type="mRNA"
            /strain="FT536"
            /db_xref="taxon:117187"
            /clone_lib="Fusarium verticillioides fcc1 mutant
            subtraction library"
            /note="Vector: pGem-T-Easy; Site_1: EcoRI; Site_2: EcoRI;
            Fungus was grown on cracked corn. RNA was isolated using
            phenol LiCl method. PolyA RNA was obtained with Oligotex
            mRNA spin columns (Qiagen). Subtracted from wild type RNA
            with PCR select cDNA subtraction kit (Clontech),
            amplified, cloned into pGem-T-EZ and transformed to E. coli
            DH5 alpha cells."
ORIGIN
Query Match      12.5%; Score 129.8; DB 6; Length 655;
Best Local Similarity 53.6%; Pred. No. 1.8e-19;
Matches 294; Conservative 0; Mismatches 252; Indels 3; Gaps 1;
QY      28 GCGCGGCGGTGACGCGCGCGGTGTTCTCGATCTCGACCGACAGTCGAGAAACG 87
DB      7 GCGCGTGTGTACCTCCGAGCGCTGAGTGTGATCTTGAGGCGGTGTCGCAAGACC 66
QY      88 ATCGGCGGTGATCGACGAGCGCGCAAGACGATGCGCCCTGATGCAATGCCAGACT 147
DB      67 ATGCACTTATCAAGAGCGCGCGCAAGCTGCTGCAAACTGTGCTTTCGCAAGTT 126
QY      148 TGGATTCCCGGTATCCCTTTTGATATGCTGCGCGCGCGCTTGAGGAGCATGCGCTTC 207
DB      127 TGGATCCCTGATATCTTACTGATGTGGAAGTCACTACTCCCAATCCTCCCATG 186
QY      208 GTCCAGCGCTATTTCGAGAAATTCGTCGTGCGCGGACGACGATGCGAGGCTTGCGG 267
DB      187 CTGAGGCGCTACCGGAGAACTCCATGCGCGCTGCACTGAGGAATGCGCGGTTCGT 246
QY      268 GATGCGCGCGCGCGCAAGCATGATGTCGCGCGGTATAGGAGCGCGCGCGCGC 327
DB      247 CGCGAGCGCGGATTAACGATCTAAGTCTTCCCTGCGCTTCTGGAATGACCAACGA 306
QY      328 AGCCTTATATGGCGGATCTTTCGCGCGCGGATGCGATCTGATGCGCGCGCGCGC 387
DB      307 ACTCTTACCTGCGCAAGTCTCATCAAGCCCGATGCGGTGATCAACGACGCTGCG 366

```

```

QY      388 AAGCTCAAGCTTACCAATCGAGCGACCGTGTTCGCGAGGAGACG---CAGCCAT 444
DB      367 AAGATCAAGCCAACTCATGTTGAGAGAGCTTTATACGCCATGTGCTCCGGGATACCTTC 426
QY      445 CTCGCGGTGACAGATACCGCATCGCGCGCTCGCGCGCTCTGTGTGCGAGACATC 504
DB      427 ATGCGCGTTAGCGAGACTGATCATCGCGCGGTGTGGCCAGCTTAACGTGCGAGAACATG 486
QY      505 CAGCATTTGTGAATATGACCATGTATGACCGCCGACGACGAGTTCACGTGCGCTCGG 564
DB      487 AACCCTTCTCAAGTCTCAACGTTTCTAGTAGAGCAGGTTACATGCTGCTTGG 546
QY      565 CCGAGCTTC 573
DB      547 CCCGCTTAC 555

RESULT 10
LOCUS      CN808140          840 bp      mRNA      linear      EST 27-MAY-2004
DEFINITION Blood_EST0223 Metarhizium anisopliae ARSEF 2575 from insect blood
            Metarhizium anisopliae cDNA clone B602 5', mRNA sequence.
ACCESSION  CN808140
VERSION     CN808140.1  GI:47729613
KEYWORDS
SOURCE      EST.
            Metarhizium anisopliae
            Metarhizium anisopliae
            Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
            Hypocreomycetidae; Hypocreales; Clavicipitaceae; mitosporic
            Clavicipitaceae; Metarhizium.
REFERENCE   1 (bases 1 to 840)
AUTHORS     Wang,C.S., Hu,G. and St. Leger,R.J.
TITLE       Gene expression profiling of Metarhizium anisopliae grown under
            different conditions: mechanisms of fungal opportunism
            unpublished (2004)
JOURNAL     Contact: Wang CS
COMMENT     Department of Entomology
            University of Maryland
            4112 Plant Sciences Building, College Park, MD 20742, USA
            Email: cwang4@umd.edu
            Seq primer: M13 Reverse.
FEATURES
source      1..840
            location/Qualifiers
            /organism="Metarhizium anisopliae"
            /mol_type="mRNA"
            /db_xref="taxon:5530"
            /clone_lib="B602"
            /note="Vector: pCMV.SPORT6.1; Metarhizium anisopliae was
            grown in insect haemolymph for 24 hours. A cDNA library
            was constructed in the vector pCMV.SPORT6.1"
ORIGIN
Query Match      12.4%; Score 129.4; DB 7; Length 840;
Best Local Similarity 53.1%; Pred. No. 2.2e-19;
Matches 299; Conservative 0; Mismatches 261; Indels 3; Gaps 1;
QY      20 AGTATCGGCGCGCGGTGACGCGCGCGGTGTTCTCGATCTCGACCGACAGTCG 79
DB      80 AGTACAAGGCTGTGTGTGTGCAAGCTGAGCTGATGCTCAACTTGAAGAAAGATCA 139
QY      80 AGAAGCGATCGGCTGATGAGAGAGCGGCGCAAGACGATGCGCTGATGCAATTC 139
DB      140 AGAAGACTGTGACTGTATCTCGAAGCTGCGAGAAAGATTGCAAGCTTATTTGCTTTC 199
QY      140 CAGAGACTTGATTCGCGGCTATCCCTTTTGATATGCTGCGCGCGCGCGCTTGCGGCA 199
DB      200 CTGAAGTTTGAATTCCTGATTTCTTATTTGCTGTGCGGTGCTCAACAAATTGCG 259
QY      200 TGCGCTTGTGCACGCTATTTCGAGAAATTCGCTGTGCGCGGAGCAAGACATGCGCAG 259
DB      260 TGCCCTCTCTCAAGAAAGTTTCAACGAAACAGCATGCGCGCGGACTCGGACGAAATGCGAC 319

```

QY 260 CCCTGGCGATGAGCGCCGCGCCAGCGCATGATGTCGTGGCCGCTATAGCGAGCG 319  
 Db 320 GAATTCGCGAGGCTGCCAAAGCCGCTAAGATCTACGTCGTGGGATATTCGAGCTCG 379  
 QY 320 CGGCGCGGACCTCTATATATGCGCGAGCGATCTTTCGCGCCCGATGCGATCTGATCGCG 379  
 Db 380 ACCGCGCATCATGTATACGTGGCGAGATCATTTGATGCCACCGGAACCGTCATCAACC 439  
 QY 380 CGGCGCGGACCTCAAGCCTTACCCTGCGAGCGGACCGCTTTCGCGAGGAGAGACGG-- 437  
 Db 440 ACCGTGCGAAGATCAAGCCCAACCATGTCAAGAACTTTGATTTGGGAGAGGAAAGCGGTG 499  
 QY 438 -CAGCCATCTCGCGGTGCAAGATACCGCATCGGCGCGCTTCGCGCGCTGTGCTGGG 496  
 Db 500 ACTCTCTTCAAGCGCGTAGTGAGACTGAATTTGGCAATCTCGCACCTGAACGTGCGG 559  
 QY 497 AGCAATCTCAGCCATTTTCGAATATAGCCCATGTAGCGCGCGAGCAAGAGTCCAGCTCG 556  
 Db 560 AAATATGAACCCCTTCTCAAGGCGCTCAACGCCAGCAAAACAGAAATCCACCTAG 619  
 QY 557 CGTGTGCGCGAGCTTCAGCCTC 579  
 Db 620 CCGCATGCGCCGTGTACGCGCCC 642

RESULT 11  
 CN808249 998 bp mRNA linear EST 27-MAY-2004  
 LOCUS Blood EST0332 Metarhizium anisopliae ARSEF 2575 from insect blood  
 DEFINITION Metarhizium anisopliae cDNA clone B711 5', mRNA sequence.  
 ACCESSION CN808249  
 VERSION CN808249.1 GI:47729722  
 SOURCE EST.  
 ORGANISM Metarhizium anisopliae  
 Metarhizium anisopliae  
 Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
 Hypocreomycetidae; Hypocreales; Clavicipitaceae; mitosporic  
 Clavicipitaceae; Metarhizium.  
 REFERENCE 1 (bases 1 to 998)  
 Wang, C.S., Hu, G., and St. Leger, R.J.  
 Gene expression profiling of Metarhizium anisopliae grown under  
 different conditions: mechanisms of fungal opportunism  
 Unpublished (2004)  
 JOURNAL Contact: Wang CS  
 COMMENT Department of Entomology  
 University of Maryland  
 4112 Plant Sciences Building, College Park, MD 20742, USA  
 Email: cswang4@umd.edu  
 Seq primer: M13 Reverse.  
 Location/Qualifiers

FEATURES  
 source  
 1..998  
 /organism="Metarhizium anisopliae"  
 /mol\_type="mRNA"  
 /db\_xref="taxon:5530"  
 /clone="B711"  
 /clone\_1lb="Metarhizium anisopliae ARSEF 2575 from insect  
 blood"  
 /note="Vector: pCMV.SPORT6.1; Metarhizium anisopliae was  
 grown in insect haemolymph for 24 hours. A cDNA library  
 was constructed in the vector pCMV.SPORT6.1"

## ORIGIN

Query Match 12.4%; Score 129.4; DB 7; Length 998;  
 Best Local Similarity 53.1%; Pred. No. 2.2e-19;  
 Matches 299; Conservative 0; Mismatches 261; Indels 3; Gaps 1;  
 QY 20 AGTATGCGCGGCGCGGCGTCAAGGCGCGGCGTCTCCGATCTCGACGCAAGTCTG 79  
 Db 104 AGTCAAGGCTGCTGTGTGCTCAAGCTGAGCTTCAACCTTGAAGAAAGCACTCA 163  
 QY 80 AGAAGGATCGGCTGTATGAGACGCGGCAAGCAAGAGAGTGTGCGCTGATCGCATTC 139

Db 164 AGAAGACTGTTCGATTCCTCGAAGCTGCGAGAAAGATTGCAAGCTTATTCCTTTC 223  
 QY 140 CAGAGACTTGATTTCCCGGCTATCCCTTTTGATATGCTGGGCGCGCGCTTGGGCA 199  
 Db 224 CTGAAGTTTGATTTCTCGATTTTCCTATTTGGCTGTGGCGTGTCACTACCAAGATTGCG 283  
 QY 200 TGGCTTTCGTCAGCGCTATTTTGAAGAAATTCGCTGTGCGCGGACAGCAAGTGGCAGG 259  
 Db 284 TGGCCCTCTCCAAAGAGTTTCAACAGAAAGCATGGCGCCGACTCGGAAGAAATGGAC 343  
 QY 260 CCCTGCGGATGCGGCGCGCGCGCCAGCGCATGATGCTGTGGCGGCTATAGGAGCGG 319  
 Db 344 GAATTCGCGAGGTGCGCAAAAGCCGCTAAGATCTACGCTGTGTGGGATATTCGAGCTCG 403  
 QY 320 CGGCGCGGACCTCTATATGAGGCGAGGATCTTCGCGCCCGAGTGGCATCTGTCGCG 379  
 Db 404 ACCGCGCATCATGTATACATGGCGAGATCATCTTATTCACCGGAACGCTATCAACC 463  
 QY 380 CGGCGCGCAAGCTCAAGCCTTACCCTATGCGGAGCGCACCGTTCGCGGAGGAGACGG-- 437  
 Db 464 ACCGTGCGAAGATCAAGCCCAACCATGTGAGAAAGCTTGTATTTGGGAGGAAAGCGGTG 523  
 QY 438 -CAGCCATCTCGCGGTGCAAGATACCGCCATCGGCGCGCTCTGTGCTGGG 496  
 Db 524 ACTCTCTTCAAGCGCGTAGTGAGACTGAATTTGGCAATCTGCGCCACTGAACGTGCGG 583  
 QY 497 AGCAATCTCAGCCATTTTCGAATATAGCCCATGTAGCGCGCGCGCAAGAGTCCAGCTCG 556  
 Db 584 AAATATGAACCCCTTCTCAAGCGGCTCAAGCCGAGCAAAACAGAAATCCACCTAG 643  
 QY 557 CGTGTGCGCGAGCTTCAGCCTC 579  
 Db 644 CCGCATGCGCCGTGTACGCGCCC 666

RESULT 12  
 DR609823 769 bp mRNA linear EST 11-JUL-2005  
 LOCUS EST999951 FvG gibberella mon11iformis cDNA clone FVGB83, mRNA  
 DEFINITION sequence.  
 ACCESSION DR609823  
 VERSION DR609823.1 GI:70684471  
 KEYWORDS EST.  
 SOURCE Gibberella mon11iformis  
 ORGANISM Gibberella mon11iformis  
 Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
 Hypocreomycetidae; Hypocreales; Nectriaceae; Gibberella.  
 REFERENCE 1 (bases 1 to 769)  
 Brown, D.W., Chung, F., Proctor, R.H., Butchko, A.E., Zheng, L., Lee, Y.,  
 Utebback, T., Smith, S., Feildlyum, F., Glenn, A.E., Plattner, R.D.,  
 Kendra, D.F., Town, C.D., and Whitefaw, C.A.  
 Analysis of 87,000 expressed sequence tags reveals alternatively  
 spliced introns in multiple genes of the fumonisin gene cluster  
 Unpublished (2005)  
 JOURNAL Contact: Brown, D.W.  
 COMMENT USDA/ARS/NCAR  
 USDA  
 1815 N. University St. Peoria, IL 61604, USA  
 Tel: 309 681 6230  
 Fax: 309 681 6689  
 Email: brownnd@ncar.usda.gov  
 TIGR sequence name: FVGB83TH  
 Seq primer: MAT TAA CCC TCA CTA AAG GG.  
 Location/Qualifiers

FEATURES  
 source  
 1..769  
 /organism="Gibberella mon11iformis"  
 /mol\_type="mRNA"  
 /strain="m1125"  
 /db\_xref="taxon:117187"  
 /clone="FVGB83"  
 /cissue\_type="mycelia"  
 /clone\_1lb="FvG"  
 /note="Vector: pBluescript II SK(+) XR; site\_1: EcoRI;



Site 2: XhoI; anamorph: *Fusarium verticillioides*. Mycelia was collected after growth in liquid GYM medium for 96 hours. Cultures were vacuum filtered and the mycelial mats were frozen in liquid nitrogen, ground to a powder, and then added to RIZOL Reagent (Invitrogen, Carlsbad CA) at approximately 1 g mycelia per 10 ml RIZOL. The cDNA was directionally ligated into the pBluescript II SK(+) XR vector (cDNA Synthesis Kit; Stratagene)."

## ORIGIN

Query Match 12.3%; Score 127.6; DB 8; Length 769;  
Best Local Similarity 51.5%; Pred. No. 5.7e-19;  
Matches 348; Conservative 0; Mismatches 319; Indels 9; Gaps 2;

```

QY 28 GGGCGGGCGGTGACAGCGCGCGCGGTCTCTCATCTCCACCGCAAGTGAAGAACG 87
DB 61 GTCGCCCCCAATCCAGAGCGAGCCGCTGGAACATCTCCAGGGTGGCGTCACAAATGCC 120
QY 88 ATCGGCTGATCGACGAGCGCGCAAGAGACGTGCGCTGATCGCATTCGAGAGACT 147
DB 121 ATCGGCTGATCGACGAGCGCGCAAGAGAGGTGTCACAGTCAATGGCTTCTGAAATC 180
QY 148 TGAATTCGCGGCTATCCCTTTGGATATGCTGGCGCGCGCGCTGGGCGATCGCTTC 207
DB 181 TTGATTCCTGATATTCATGAGCATCTGGGCCAACTCGCTTACCGAAGCGACCATGG 240
QY 208 GTCCAGCGCTATTTGGAATTCGCTGCTGGCGCGCGCAAGACAGTGGCAGGCGCTTC 267
DB 241 ATCAATAGTACTTCAAGAACTCAATGAGAAAGTACCTGAGTATGACCAATCCGA 300
QY 268 GATGCGCGCGCGCGCGCATGCTGCTGGCGCGCTATGACGAGCGCGCGCGCGC 327
DB 301 GCTGCTTTGAGAGGAGGAGGTCTTTTGTATCTTGAATACGAGAGATACAGGAGGA 360
QY 328 AGCTCTATATGAGCGCGCATCTTCGCGCGCGCGCATGATGCGCGCGCGCGC 387
DB 361 ACCCTTATATGAGCGAGTCTTTTATGAGAGAGCGCGCATATTTGTTTCCACCTGCG 420
QY 388 AAGCTCAAGCTTACCTTACGCGAGCGACCGTGTTCGCGAGGAGAACCGGACCATCT- 446
DB 421 AAGATCAAGCTTACCTTACGCGAGCGTGTATCTTACGCTGACGAGGAGGAGTCTCTG 480
QY 447 --GGCGGTGACGATACCGCATGCGGCGCGCTGCGCGCGCTGCTGTGCGAGACATC 504
DB 481 ACCTTATATGAGCGCGCATCTTCGCGCGCGCGCATGATGCGCGCGCGCGC 540
QY 505 CAGCATTTGTAATATGCGCATGTACGCGCGCGCAAGACAGTGCAGTGCCTGCG 564
DB 541 CAGGACATTTCCGCTACTAGATATCTCCAGAGACGTGATTCACAGTCTCCAGCTGG 600
QY 565 CCGAGCTTCAAGCTCTA-----TGGCGCATGCGCTATGCGCTGCGAGGTCAT 618
DB 601 CTTTATCTTCCCGCGAAGAGTCCCGAGTGGCATACCAATCATCTCCGAATGTGCG 660
QY 619 ACGCGCGCAAGCAGATCTACGCGGTGAGGCGCGCTGATACGTGCTGGGCTGCTGCGG 678
DB 661 AAGGCTTCTCTCAAGTGTCTCATGAGAGAGAGCTGCTTCTGCTGCGCAAGTCA 720
QY 679 ACCGTTTCCGCGGAGA 694
DB 721 ATCATGACTGAGAGAGA 736

```

RESULT 13  
LOCUS DR709497 840 bp mRNA linear EST 14-JUL-2005  
DEFINITION *Aspergillus niger* pBluescript (EcoRI-XhoI) *Aspergillus*  
niger cDNA clone *Asn\_11063*, mRNA sequence.

ACCESSION DR709497  
VERSION DR709497.1 GI:70825788  
KEYWORDS EST.  
SOURCE *Aspergillus niger*  
ORGANISM *Aspergillus niger*

REFERENCE  
AUTHORS Tsang A., Storms R. and Bulter G.  
TITLE Expressed sequence tags from *Aspergillus niger* cDNA library  
JOURNAL Unpublished (2005)  
COMMENT Contact: Tsang A  
Centre for Structural and Functional Genomics  
Concordia University  
7141 Sherbrooke Street West, SP545-1, Montreal, QC H4B1R6, Canada  
Tel: 514 848 2424 3405  
Fax: 514 848 4504  
Email: tsang@vax2.concordia.ca  
POLYA=No.

## FEATURES

## source

1. 840  
location/Qualifiers

/organism="Aspergillus niger"  
/mol\_type="mRNA"  
/strain="N402"  
/db\_xref="taxon:5061"  
/clone="Asn\_11063"  
/dev\_stage="mycelial growth"  
/lab\_host="E. coli"  
/note="Vector: pBluescript KS+; Site 1: XhoI; Site 2: EcoRI; Complementary DNA was synthesized with ZAP kit (Stratagene) using poly(A)+RNA isolated from *Aspergillus* niger cultured under different carbon sources (glucose, maltose, xylose, lactose, sorbitol, xylan, and bran). Synthesis was primed with oligo(dT)/XhoI primer. EcoRI adaptors were ligated to the blunt-ended, double-stranded cDNA. The EcoRI-XhoI-digested cDNA was ligated with EcoRI-XhoI-digested pBluescript KS+ (Invitrogen Corp)."

## ORIGIN

Query Match 11.8%; Score 122.6; DB 8; Length 840;  
Best Local Similarity 50.3%; Pred. No. 8.2e-18;  
Matches 417; Conservative 0; Mismatches 394; Indels 18; Gaps 4;

```

QY 25 CGCGCGCGCGGTGACAGCGCGCGCGGTCTCTCATCTCCACCGCAAGTGAAGAA 84
DB 19 CGCGTGCAGTACCCAGCGAGAGCCAGTCTGCTGATGGAAGCAACCGTGAAGAA 78
QY 85 GCATGCGCTGATTCGAGCGCGCGCGCAAGACAGTGTGCGCTGATTCGATTCGAG 144
DB 79 ACATGTACCTTATTTGTGTAAGCAGTGCAGACGGCGCTGATGATTTTCCGAG 138
QY 145 ACTTGATTCGCGGCTATCCCTTTGGATATGCTGGCGCGCGCTGCGGCGCGC 204
DB 139 TGTGGATCCGGGATATCTGCTGATTTGG-----GACGCGCTGTGACATGCGC 192
QY 205 TTGTCAGCGCTATTTGGAATTCGCTGCTGCGCGCGCAAGACAGTGGAGCCCTG 264
DB 193 CTATCATCTATTTAATACAAAATCCCTGAATGATTCCTCCGAATATGGCAGATC 252
QY 265 GCGGATGCGCGCGCGCGCAAGCAGTGTGCTGCGCTATTAAGCGCGCGCGG 324
DB 253 CAGCAATGCGCAAGAGATTAATATGCTGTGTGTGTGCGCTTTTCCGAAACCTGCAT 312
QY 325 GCGAGCTCTATATGAGCGAGCGATCTTGGCGCGCGCGAGCATGATGCGCGCGG 384
DB 313 AACTCTCTATATCTCGAGCGCTATTTATTTGCAAGTATGGAAGATCTTACACACCGC 372
QY 385 CGCAAGCTCAAGCTTACCAATGCGAGCGCACCGTGTTCGCGA--GGAGACGCGAGC 441
DB 373 AAGAAATCAAGCCACTATATGAGCGGACCATTTTGGTGATCTGTTGAGACTGT 432
QY 442 CATCTGCGGTGACGATACCGCATTCGCGCGCTGCGCGCGCTCTGTTCTGGAGAC 501
DB 433 CTGCAAGTGTATGATCATCATCAGCGCGGTGTGTGTCTTTCTCTGCTGGAGCAT 492
QY 502 ATCAGCATTTGGAATATGCGCATGTACGCGCGCGAGCAAGGTCACGTGCGGTG 561

```

Db 493 ATCCAGCCTCTGCTCAAGTATCACTTATGCGAGCGGGAACAAATTCATGTTGCCGA 552  
 QY 562 TGGCCGAGCTTACAGCTCTATTCGCGAGTGGCTTA-----TGGCTCGAGACCGAGATC 615  
 Db 553 TGGCCGCTCTGTTCTCTCATAGGAGAGCGTTCTTGTGTTGATGTCCACCGAGGCG 612  
 QY 616 AATAACGCCGACAGCCAGATCTACGCGGTGAGAGGCGGCTGCTACGTCTGGCGTCTGC 675  
 Db 613 ACCAGTTTCATTGCGAGGACCTATGCGAAGATCAAGTCTTTTGTCTTCATAC--C 669  
 QY 676 GCGACCGTTTCGCGGAGATGATCAAGTATGTTGTGATACGCCCGCAAGAGATGTTTC 735  
 Db 670 ACAACAGTCAATAGGCGCAGTCTGGCATTTGACCGAGTGGCAATGAGCTGGGGCCCTGATG 729  
 QY 736 CTGAAGCGCGCGCGCGGTTTTCATGATTTTGGGGCCGAGCGCGCGCTGGCGGAG 795  
 Db 730 AGCACCCCTGAGGAGTCTCTCGCATCTTGGGCGCGACGCTGCCAATTATGCAA 789  
 QY 796 CCGCTCCCGGAGACCGAAGAGGAGCTGCTGCTGCGCCGATATCGACCTCG 844  
 Db 790 CCATTCCTATATGAGAGAGGAGGATTTATCTACCCGACCTCGACTTTG 838

RESULT 14  
 LOCUS A0160684 580 bp DNA linear GSS 09-SEP-1998  
 DEFINITION mgxb0006104r CUGI Rice Blast BAC Library Magnaporthe grisea genomic  
 clone mgxb0006104r, genomic survey sequence.  
 VERSION A0160684  
 KEYWORDS A0160684.1 GI:3557673

SOURCE Magnaporthe grisea (anamorph: Pyricularia grisea)  
 ORGANISM Magnaporthe grisea  
 Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
 Sordariomycetes; incertae sedis; Magnaportheaceae; Magnaporthe.  
 1 (bases 1 to 580)  
 Yu, Y., Zhu, H., Boyd, C.A., Gaudette, B., Gayle, A., Kingsbury, R.,  
 Phillips, K., Sasinowski, M., Wing, R.A. and Dean, R.A.  
 A BAC End Sequencing Framework to Sequence the Magnaporthe grisea  
 Genome

JOURNAL Unpublished (1998)  
 COMMENT Contact: Dean RA  
 Clemson University Genomics Institute  
 Clemson University  
 100 Jordan Hall, Clemson University, Clemson, SC 29634  
 Tel: 864 656 5737  
 Fax: 864 656 4293  
 Email: rdean@clemson.edu  
 Seq primer: GGAACAGCTATGACCATG  
 Class: BAC ends  
 High quality sequence stop: 357.

FEATURES  
 source location/Qualifiers

1..580  
 /organism="Magnaporthe grisea"  
 /mol\_type="genomic DNA"  
 /strain="70-15"  
 /db\_xref="taxon:148305"  
 /clone="mgxb0006104r"  
 /issue\_type="Protoplasts"  
 /lab\_host="E. coli DH10B"  
 /clone\_1ib="CUGI Rice Blast BAC Library"  
 /note="Vector: PBACWICH; Site\_1: HindIII; Site\_2: HindIII;  
 Rice Blast is one of the most devastating fungal diseases  
 of rice world wide. It is a filamentous ascomycete with  
 a haploid genome (n=7) of approximately 40 MBP. Rice  
 blast is an important model fungal pathogen for studying  
 numerous aspects of the fungal-host interaction. In  
 order to facilitate genome wide analysis, a BAC library  
 containing 9216 clones with an average insert size of 130  
 kbp was constructed. This library represents greater  
 than 25X genome coverage. High density colony filters  
 are available upon request."

ORIGIN

Query Match 11.2%; Score 116.6; DB 9; Length 580;  
 Best Local Similarity 56.2%; Pred. No. 2e-16;  
 Matches 240; Conservative 0; Mismatches 184; Indels 3; Gaps 1;

QY 143 AGACTTGATATCCCGCTATCCCTTTTGATATGCTGGGCGCGCGCTGGGCGATGC 202  
 Db 82 AAACAGGATCTTGGCTACCCCTATTTGATGTAAGTGAACATCTTGCACAGCTCC 141  
 QY 203 GCTTGTCCAGCCGCTATTTTGAGAAATTCGCTGCGCGCGCAGCAGAGTGCAGGCC 262  
 Db 142 CGATGCTCAAGGCTACCGCGAGAACCTCTCAAGTTGACTGGAAGATGCGCGCA 201  
 QY 263 TGGCGATGGCGCGCGCGCGCGCAGCGATGATGATCTGTCGCGCGCTATAGAGCGCGG 322  
 Db 202 TCCGGGCGGCGGCTCCGACACAAACAAATCTACGCTCACTGAGGCTTCTCGAATTTGACC 261  
 QY 323 GCGGCAAGCTCTATATGAGCGCAGCGATCTTGGCGCGCGCGCGATGCGCGCGC 382  
 Db 262 AGCGAACCTGTATCTGGGCGAGGTGCTGATGGCGCGCGCGAGGCTCAATCAC 321  
 QY 383 GCGGCAAGCTCAAGCCTTACATGCGGAGCGCACCGTTCGCGCGGAGAGAGCG--CA 439  
 Db 322 GCGGCAAGATCAAGCCAAACCATGTCGAGAAAGCTTGTATGCGGATGCTCGGATGA 381  
 QY 440 GCCATCTCGGCGGTGACGATACCGCATCGGCGCGCTCGCGCGCTGTTGCTGGAGC 499  
 Db 382 GCTTCATACCCGTGAGCGAAACAAACACTCGGCGCGCTCGGTTCAACTGCTGGAGAG 441  
 QY 500 ACATTCAGCATTTGTGAATATACGATGATGAGCGCGCGCGACGACAGGTTCACGTGCGT 559  
 Db 442 ACATGAATCGGCTCCAAAGTGTGATGAGCGTGGCGCGGTGCGAGGTCCATGTCGCG 501  
 QY 560 CGTGCC 566  
 Db 502 CGTGCC 508

RESULT 15  
 LOCUS DR709485 804 bp mRNA linear EST 14-JUL-2005  
 DEFINITION Asn\_11049 Aspergillus niger pBluescript (EcoRI-XhoI) Aspergillus  
 niger cDNA clone Asn\_11049, mRNA sequence.  
 VERSION DR709485  
 KEYWORDS DR709485.1 GI:70825776

SOURCE Aspergillus niger  
 ORGANISM Aspergillus niger  
 Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
 Eurotiiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.  
 1 (bases 1 to 804)  
 Tsang, A., Storms, R. and Bulter, G.  
 Unpublished (2005)  
 Title: Expressed sequence tags from Aspergillus niger cDNA library  
 JOURNAL Unpublished (2005)  
 COMMENT Contact: Tsang A  
 Centre for Structural and Functional Genomics  
 Concordia University  
 7141 Sherbrooke Street West, SP545-1, Montreal, QC H4B1R6, Canada  
 Tel: 514 848 2424 3405  
 Fax: 514 848 4504  
 Email: tsang@vax2.concordia.ca  
 POLYA=No.

FEATURES  
 source location/Qualifiers

1..804  
 /organism="Aspergillus niger"  
 /mol\_type="mRNA"  
 /strain="NA402"  
 /db\_xref="taxon:5061"  
 /clone="Asn\_11049"  
 /dev\_stage="mycelial growth"  
 /lab\_host="E. coli"  
 /clone\_1ib="Aspergillus niger pBluescript (EcoRI-XhoI)"  
 /note="Vector: pBluescript KS+, site\_1: XhoI; site\_2:



**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 27, 2006, 00:01:50 ; Search time 5618.87 Seconds  
(Without alignments)  
10531.310 Million cell updates/sec

Title: US-09-751-299-1

Perfect score: 1041  
Sequence: 1 atgtcgagagccatgacgaa.....gcgcgcgcgcgcgcagtag 1041

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 5683141 seqs, 2842172563 residues

Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database :

GenBml:1:  
1: gb\_ba:\*  
2: gb\_in:\*  
3: gb\_env:\*  
4: gb\_cm:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pr:\*  
9: gb\_ro:\*  
10: gb\_stra:\*  
11: gb\_sy:\*  
12: gb\_un:\*  
13: gb\_vl:\*  
14: gb\_hlg:\*  
15: gb\_pl:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query	Length	DB	ID	Description
1	1041	100.0	1041	3	AY487497	AY487497 Unculture
2	1041	100.0	1041	6	AX189644	AX189644 Sequence
3	682	65.5	1026	3	AY487438	AY487438 Unculture
4	486	46.7	1062	3	AY487543	AY487543 Unculture
5	457.2	43.9	1065	3	AY487496	AY487496 Unculture
6	455.2	43.7	1077	3	AY487522	AY487522 Unculture
7	454.6	43.7	1017	3	AY487536	AY487536 Unculture
8	453.6	43.6	1017	3	AY487558	AY487558 Unculture
9	450.4	43.3	1017	3	AY487475	AY487475 Unculture
10	448.6	43.2	1005	3	AY487559	AY487559 Unculture
11	447.8	43.1	1047	3	AY487430	AY487430 Unculture
12	447.8	43.0	1062	3	AY487552	AY487552 Unculture
13	445	42.7	110000	1	AP006618_34	Continuation (35 o
14	440.2	42.3	1017	3	AY487448	AY487448 Unculture
15	438	42.1	1017	3	AY487436	AY487436 Unculture
16	437	42.0	1017	3	AY487531	AY487531 Unculture
17	436	41.9	1014	3	AY487464	AY487464 Unculture
18	427.2	41.0	1017	3	AY487480	AY487480 Unculture

19	423.2	40.7	984	3	AY487523	AY487523 Unculture
20	422.8	40.6	1014	3	AY487429	AY487429 Unculture
21	422	40.5	1038	3	AY487427	AY487427 Unculture
22	420	40.3	1014	3	AY487463	AY487463 Unculture
23	408.2	39.2	105027	1	AJ617740	AJ617740 Pseudomon
24	399.6	38.4	1059	6	CQ874190	CQ874190 Sequence
25	398	38.2	6976	1	AY885240	AY885240 Pseudomon
26	388.8	37.3	1038	3	AY487470	AY487470 Unculture
27	385.2	37.0	1014	3	AY487479	AY487479 Unculture
28	382.2	36.7	1008	3	AY487545	AY487545 Unculture
29	381	36.6	1014	3	AY487515	AY487515 Unculture
30	380.4	36.5	110000	1	CP000075_00	CP000075 Pseudomon
31	379.6	36.5	1059	3	AY487532	AY487532 Unculture
32	377.2	36.2	1011	3	AY487546	AY487546 Unculture
33	376.8	36.2	110000	1	AE016853_02	Continuation (3 of
34	370.6	35.6	1041	3	AY487481	AY487481 Unculture
35	370.2	35.6	1014	3	AY487445	AY487445 Unculture
36	370.2	35.6	1014	3	AY487449	AY487449 Unculture
37	353.2	33.9	1050	3	AY487432	AY487432 Unculture
38	352.8	33.9	1014	3	AY487473	AY487473 Unculture
39	345.4	33.2	1002	3	AY487492	AY487492 Unculture
40	345.2	33.2	999	3	AY487512	AY487512 Unculture
41	343.2	33.0	1035	6	AR036234	AR036234 Sequence
42	343.2	33.0	1200	6	AR036235	AR036235 Sequence
43	343.2	33.0	1200	6	E12616	E12616 DNA encodin
44	341.2	32.8	1011	3	AY487507	AY487507 Unculture
45	328.2	31.5	110000	1	BA000040_70	Continuation (71 o

#### ALIGNMENTS

RESULT 1	AY487497	1041 bp	DNA	linear	ENV 05-APR-2004
LOCUS	AY487497	Uncultured organism clone 2A13	nitrilase (BD5086)	gene, complete	
DEFINITION	AY487497	Uncultured organism clone 2A13	nitrilase (BD5086)	gene, complete	
ACCESSION	AY487497	GI:40890198			
VERSION	AY487497.1				
KEYWORDS	ENV.				
SOURCE	uncultured organism				
ORGANISM	unclassified; environmental samples.				
REFERENCE	1 (bases 1 to 1041)				
AUTHORS	Robertson,D.E., Chaplin,J.A., Desantis,G., Podar,M., Madden,M., Chi,E., Richardson,T., Milan,A., Miller,M., Weiner,D.P., Wong,K., McQuaid,J., Farwell,B., Preston,L.A., Tan,X., Sneed,M.A., Keller,M., Mathur,E., Kretz,P.L., Burk,M.J. and Short,J.M.				
TITLE	Exploring nitrilase sequence space for enantioselective catalysis				
JOURNAL	Appl. Environ. Microbiol. 70 (4), 2429-2436 (2004)				
REFERENCE	2 (bases 1 to 1041)				
AUTHORS	Robertson,D.E., Chaplin,J.A., Desantis,G., Podar,M., Madden,M., Chi,E., Richardson,T.H., Milan,A., Miller,M., Weiner,D.P., Wong,K., McQuaid,J., Farwell,B., Preston,L.A., Tan,X., Keller,M., Mathur,E., Kretz,P.L., Burk,M.J. and Short,J.M.				
TITLE	Direct Submission				
JOURNAL	Submitted (26-NOV-2003) Bioinformatics, Diversa Corporation, 4955				
FEATURES	Location/Qualifiers				
source	1..1041				
	/organism="uncultured organism"				
	/mol_type="genomic DNA"				
	/db_xref="taxon:155900"				
	/clone="2A13"				
	/environmental_sample				
gene	1..1041				
	/locus_tag="BD5086"				
CDS	1..1041				
	/locus_tag="BD5086"				
	/EC_number="3.5.5.7"				
	/codon_start=1				
	/evidence="experimental"				
	/product="nitrilase"				

[illegible]

ORIGIN

Query Match	100.0%;	Score 1041;	DB 3;	Length 1041;
Best Local Similarity	100.0%;	Pred. No. 5.1e-124;		
Matches 1041;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

QY	1	ATGTCCGAGGCCATGACGCAAGTATTCGGGGCGGGCGGTGACAGGGCCGCGCGGTGTTCTC	60
Db	1	ATGTCCGAGGCCATGACGCAAGTATTCGGGGCGGGCGGTGACAGGGCCGCGCGGTGTTCTC	60
QY	61	GATCTCCAGCCGACAGTCCGAGAAAACGATCGGCCTGATCCAGCAGCGGCCCAAGCAGAC	120
Db	61	GATCTCCAGCCGACAGTCCGAGAAAACGATCGGCCTGATCCAGCAGCGGCCCAAGCAGAC	120
QY	121	GTGGGCGCTGATCGGATTCGCCAGAGCTTGAAATCCCGGCTATCCCTTTTGGATATAGCTG	180
Db	121	GTGGGCGCTGATCGGATTCGCCAGAGCTTGAAATCCCGGCTATCCCTTTTGGATATAGCTG	180
QY	181	GGCGCGCCGCGCTTTGGGGCAGTCGCTTCGTCCAGCGCTAATTTCCGAATTTGCTCGTGGC	240
Db	181	GGCGCGCCGCGCTTTGGGGCAGTCGCTTCGTCCAGCGCTAATTTCCGAATTTGCTCGTGGC	240
QY	241	GGCAGCAAGCAGTGGCAGAGCCCTGGCGGATACCGGCGCGCCGACACGGATGATGTCGTG	300
Db	241	GGCAGCAAGCAGTGGCAGAGCCCTGGCGGATACCGGCGCGCCGACACGGATGATGTCGTG	300
QY	301	GCCGCGCTATAGCAGCGCGCGGGCGGCGAGCCTCTATATAGGAGCAAGCGCATTTGCGCCCC	360
Db	301	GCCGCGCTATAGCAGCGCGCGGGCGGCGAGCCTCTATATAGGAGCAAGCGCATTTGCGCCCC	360
QY	361	GATGGCGCATCTGATTCGCGCGCGCGCGCGGAAAGTTCAAGCCTTACCATTGCGGAGCGACCGTG	420
Db	361	GATGGCGCATCTGATTCGCGCGCGCGCGCGGAAAGTTCAAGCCTTACCATTGCGGAGCGACCGTG	420
QY	421	TTCCGCGAGGAGAGACGCGCAGGCATCTCGCGGTGACAGATACCGGCATTCGGCGCTCGGC	480
Db	421	TTCCGCGAGGAGAGACGCGCAGGCATCTCGCGGTGACAGATACCGGCATTCGGCGCTCGGC	480
QY	481	GGCGCTGTTGTCCTGGGAGCAATCCAGGCAATTTGTCGAAATACCGCATGTAACCGCGCGAC	540
Db	481	GGCGCTGTTGTCCTGGGAGCAATCCAGGCAATTTGTCGAAATACCGCATGTAACCGCGCGAC	540
QY	541	GACAGGTCACGTCGCGTCGTGGCCGAGTTTCAGCCTCTATTCGCGGATGAGCTAATGCG	600
Db	541	GACAGGTCACGTCGCGTCGTGGCCGAGTTTCAGCCTCTATTCGCGGATGAGCTAATGCG	600
QY	601	CTCGGACCCGAGGTCATATCCGCGCGCAAGCCAGATCTACGCGGTGCAAGGCGGCTGCTAC	660
Db	601	CTCGGACCCGAGGTCATATCCGCGCGCAAGCCAGATCTACGCGGTGCAAGGCGGCTGCTAC	660
QY	661	GTGCTGGCGTTCGTCGCGGACCGTTTTCGCGGAGATGATCAAGGTAATTGGTGGATACGCC	720
Db	661	GTGCTGGCGTTCGTCGCGGACCGTTTTCGCGGAGATGATCAAGGTAATTGGTGGATACGCC	720
QY	721	GACAAAGAGATGTTCTCTCAAGGCGCGCGCGGCTTTTGCAATGATTTTCGGGCCCGACGGC	780
Db	721	GACAAAGAGATGTTCTCTCAAGGCGCGCGCGGCTTTTGCAATGATTTTCGGGCCCGACGGC	780
QY	781	CGCGCCTCTGGCGGAGCGCTCCCGGAGACCGAAAGGGACATGCTGTCGTCGCGATATCGAC	840
Db	781	CGCGCCTCTGGCGGAGCGCTCCCGGAGACCGAAAGGGACATGCTGTCGTCGCGATATCGAC	840
QY	841	CTCGGCATGATCGCGTTGGCCAGGCGCGGCGCAATCCGCGGCGCACTATTCACGCGCC	900
Db	841	CTCGGCATGATCGCGTTGGCCAGGCGCGGCGCAATCCGCGGCGCACTATTCACGCGCC	900

Db	841	CTCGGATGATCGGTTTGGCCAAAGCGGCGGCGATTCGGGGGCGACTATTTCACGGGCC	900
QY	901	GACGTAACGCGCGGCTGCTGTGGATCGACGTCGCCGCCAACGCGTCTGTCACGCTTGATGCC	960
Db	901	GACGTAACGCGCGCTGCTGTGGATCGACGTCGCCGCCAACGCGTCTGTCACGCTTGATGCC	960
QY	961	GCATTCCGAACCGCAAAACGAGGACAAAGGCGACGCGCGCTGCGCGCTGTGTGGCGGAA	1020
Db	961	GCATTCCGAACCGCAAAACGAGGACAAAGGCGACGCGCGCTGCGCGCTGTGTGGCGGAA	1020
QY	1021	AGCGCGCGCGCGCGCAGTAG	1041
Db	1021	AGCGCGCGCGCGCGCAGTAG	1041

RESULT 2				
AX189644	AX189644	1041 bp	DNA	linear
LOCUS	AX189644			PAT 08-AUG-2001
DEFINITION	Sequence 1 from Patent WO0148175.			
ACCESSION	AX189644			
VERSION	AX189644.1	GI:15143034		
KEYWORDS				
SOURCE	unidentified			
ORGANISM	unidentified			
	unclassified sequences.			

AUTHORS	Madden, M.D. and Chaplin, J.A.
TITLE	Methods for producing enantiomerically pure $\gamma$ (a)-substituted
JOURNAL	carboxylic acids
	Patent: WO 0148175-A 1 05-JUL-2001;
	DIVERSA CORPORATION (US)

FEATURES	source	Location/Qualifiers
	1. .1041	/organism="unidentified"
		/mol_type="unassigned DNA"
		/db_xref="taxon:32644"
		/note="Description of Unknown Organism: Obtained from an environmental sa mpe"
CDS	1. .1041	

```

/note="unnamed protein product"
/codon_start=1
/protein_id="CAC50776.1"
/db_xref="GI:15143035"
/translation="MSBPMKRVGAANVOAFLPLDLRTVEKAIGLLEOAKKDVRLLI
APRWTPGYPFWITGAPAKMGKRFUORFVSELTIGSKCOMKALADRRHWGQ
YSEBAGSLTMCQILFGPDDDLIARRKLPHTAERIVFEESGSHLAVDYTLGRG
ALCWEHETQFLSKTAMTAAEQVAVASVPSFSLRGKVALGFEVNTASQITAVEEG
CYVALSCATVSPENITKVLVTPDKEMFLTKGGSGPAMI FGDGSALEPFLDETEGEL
ADDIDGMI ALAKAAPAGHSRDPDVTIRLLDDRPAORVVTLLAAEPQNEEDGDAPA
LRVAESFAAAQ"

```

Query Match	100.0%	Score 1041	DB 6	Length 1041
Best Local Similarity	100.0%	Pred. No. 5.1e-124	Mismatches 0	Indels 0
Matches 1041	Conservative	0	Mismatches 0	Gaps 0
QY	1	ATGTGGAGGCCATGACGAAGTATCGGGCGCGCGGTGCAGGCCCGCGGTGTTCTC	60	
DB	1	ATGTGGAGGCCATGACGAAGTATCGGGCGCGCGGTGCAGGCCCGCGGTGTTCTC	60	
QY	61	GATTCGACCGGCACAGTCGAGGAAGGCAATCGGCTGTATCGAGCAGGGCGCCAGCGAGAC	120	
DB	61	GATTCGACCGGCACAGTCGAGGAAGGCAATCGGCTGTATCGAGCAGGGCGCCAGCGAGAC	120	
QY	121	GTGGCGCTGATTCGCATTCCAGAGACTTGGATTCCCGGCTATCCCTTTTGGATATGGCTG	180	
DB	121	GTGGCGCTGATTCGCATTCCAGAGACTTGGATTCCCGGCTATCCCTTTTGGATATGGCTG	180	
QY	181	GGCGCGCGCGCTTGGGGGCATGCGCTTGTCTCAGCGCATTTTCAGAAATTCGCTCGGCG	240	
DB	181	GGCGCGCGCGCTTGGGGGCATGCGCTTGTCTCAGCGCATTTTCAGAAATTCGCTCGGCG	240	
QY	241	GGCGGCAAGCAGTGGCAGGCGCTTGGCGGATGCGGCCCGCGCCACAGGACGATGTCGTG	300	

Db 241 GGCGAAGACGATGGCGAGCCCTGGCGATCGGCGCCGCCACGCGATCACTGTG 300  
Qy 301 GCCCGATTATAGCGAGCGCGCGCGCGAGCCTTATATAGGCGACGCGATTTGCGCC 360  
Db 301 GCCCGATTATAGCGAGCGCGCGCGCGAGCCTTATATAGGCGACGCGATTTGCGCC 360  
Qy 361 GATGGCATCTGATCGCGCGCGCGCGAGCTCAAGCTTCAAGCTTCAAGCTTCAAGCTT 420  
Db 361 GATGGCATCTGATCGCGCGCGCGCGAGCTTCAAGCTTCAAGCTTCAAGCTTCAAGCTT 420  
Qy 421 TTCCGCGAGGAGAGCGCGAGCCATCTCGCGGTGACGATACCGCATCGCGCGCTGCG 480  
Db 421 TTCCGCGAGGAGAGCGCGAGCCATCTCGCGGTGACGATACCGCATCGCGCGCTGCG 480  
Qy 481 GCGCTCTGTTGCTGGAGACATCAAGCATTTGCGAAATAGCCCATGATGCGCGCGAC 540  
Db 481 GCGCTCTGTTGCTGGAGACATCAAGCATTTGCGAAATAGCCCATGATGCGCGCGAC 540  
Qy 541 GAACAGGTCCACGTCGCGGTGCGTGGCGAGCTTCAAGCTTCAAGCTTCAAGCTTCAAG 600  
Db 541 GAACAGGTCCACGTCGCGGTGCGTGGCGAGCTTCAAGCTTCAAGCTTCAAGCTTCAAG 600  
Qy 601 CTGCGACCGGAGGTCAATACCGCGCAAGCCAGATCTACGCGGTGCGAGGCGCGCTGAC 660  
Db 601 CTGCGACCGGAGGTCAATACCGCGCAAGCCAGATCTACGCGGTGCGAGGCGCGCTGAC 660  
Qy 661 GTGCTGCGGTGCG 720  
Db 661 GTGCTGCGGTGCG 720  
Qy 721 GACAGAGATGTTCTTCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 780  
Db 721 GACAGAGATGTTCTTCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 780  
Qy 781 CG 840  
Db 781 CG 840  
Qy 841 CTGCGACGTATCGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 900  
Db 841 CTGCGACGTATCGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 900  
Qy 901 GACCTACCG 960  
Db 901 GACCTACCG 960  
Qy 961 GCATTGCAACCGCAAAAGAGAGCAAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1020  
Db 961 GCATTGCAACCGCAAAAGAGAGCAAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1020  
Qy 1021 AGCG 1041  
Db 1021 AGCG 1041

RESULT 3  
AY487438 1026 bp DNA linear ENV 05-APR-2004  
LOCUS Uncultured organism clone 2A14 nitrilase (BD7266) gene, complete  
DEFINITION cds  
ACCESSION AY487438  
VERSION AY487438.1 GI:40890080  
KEYWORDS ENV  
SOURCE uncultured organism  
ORGANISM uncultured organism  
unclassified; environmental samples.  
REFERENCE 1 (bases 1 to 1026)  
AUTHORS Roberton,D.E., Chaplin,J.A., Desantis,G., Podar,M., Madden,M.,  
Chl,E., Richardson,T.H., Milan,A., Miller,M., Weiner,D.P., Wong,K.,  
McQuaid,J., Farwell,B., Preston,L.A., Tan,X., Snad,M.A.,  
Keller,M., Mathur,E., Kretz,P.L., Burk,M.J. and Short,J.M.  
Exploring nitrilase sequence space for enantioselective catalysis

JOURNAL Appl. Environ. Microbiol. 70 (4), 2429-2436 (2004)  
REFERENCE 2 (bases 1 to 1026)  
AUTHORS Roberton,D.E., Chaplin,J.A., Desantis,G., Podar,M., Madden,M.,  
Chl,E., Richardson,T.H., Milan,A., Miller,M., Weiner,D.P., Wong,K.,  
McQuaid,J., Farwell,B., Preston,L.A., Tan,X., Snad,M.A.,  
Kretz,P.L., Burk,M.J. and Short,J.M.  
TITLE Direct Submission  
JOURNAL Submitted (26-NOV-2003) Bioinformatics, Diversa Corporation, 4955  
Directors Place, San Diego, CA 92121, USA  
FEATURES  
source  
1..1026  
/organism="uncultured organism"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:155900"  
/clone="2A14"  
/environmental\_sample  
1..1026  
/locus\_tag="BD7266"  
/locus\_tag="BD7266"  
/EC\_number="3.5.5.7"  
/evidence="experimental"  
/product="nitrilase"  
/protein\_id="FAR97385.1"  
/db\_xref="GI:40890081"  
/translation="MISPVTVYRAAQAQAPSLDIDRYEKTIALIEQAEDPRLI  
AFPEPTIPEGPLMIVGSPAMGRFVORFENSLVSGKQWNAIADAARHRTVVVG  
FSERAGSLYMGQALFGBELIAARBLPTFAERTVFGESGSHLAVETVGRIG  
ALCSWHIOLPSKYAMVANEQVHWSMPCFSLYRGVAYLGEVNTAASQVAVEGG  
CYLAASGLVTPPELILVLDTPDKBRLILAGGFSFMI FGPDRALALQPLETEGLVT  
AEIDDAIALAKAADPAGHYARPDVTRLINRPARVADALPREVVGSEDAEPPT  
QPEAAD"

ORIGIN  
Query Match 65.5%; Score 682; DB 3; Length 1026;  
Best Local Similarity 79.4%; Pred. No. 4e-78;  
Matches 808; Conservative 0; Mismatches 210; Indels 0; Gaps 0;  
Qy 1 ATGTGAGAGCCATGACGAAGTATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 60  
Db 1 ATGTAAAGTCCCGTGAAGCGATATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 60  
Qy 61 GATCTGACCGCGCACGTCGAGAAAGCATGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 120  
Db 61 GATCTGACCGCGCACGTCGAGAAAGCATGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 120  
Qy 121 GTGCGCGTATGCGCATTTCCGAGAGCTTGATTCGCGCGCGCGCGCGCGCGCGCGCG 180  
Db 121 GTGCGCGTATGCGCATTTCCGAGAACTTGATTCGCGCGCGCGCGCGCGCGCGCGCG 180  
Qy 181 GCG 240  
Db 181 GCGTGGCG 240  
Qy 241 GCGAGCAGAGGTGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 300  
Db 241 GCGAGCAGAGGTGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 300  
Qy 301 GCGCGCATATGCGAGCG 360  
Db 301 GCGCGCATATGCGAGCG 360  
Qy 361 GATGGCATCTGATCG 420  
Db 361 GATGGCATCTGATCG 420  
Qy 421 TTCCGCGAGGAGAGCGCGAGCCATCTCGCGGTGACGATACCGCATCGCGCGCTGCG 480  
Db 421 TTCCGCGAGGAGAGCGCGAGCCATCTCGCGGTGACGATACCGCATCGCGCGCTGCG 480  
Qy 481 GCGCTCTGTTGCTGGAGACATCAAGCATTTGCGAAATAGCCCATGATGCGCGCGAC 540  
Db 481 GCGCTCTGTTGCTGGAGACATCAAGCATTTGCGAAATAGCCCATGATGCGCGCGAC 540

Db 481 GCCCTCTGCTGAGACATCCAGCCCTCTCGAAATACGAGATGATGCGGCAAC 540  
QY 541 GAACAGTCCAGCTGCGCTGCGCCGAGCTTTCAGCCTCTATCCGCGATGCTATGCG 600  
Db 541 GAACAGTCCAGCTGCGCTGCGCCGAGCTTTCAGCCTCTATCCGCGATGCTATGCG 600  
QY 601 CTGCGACCGGAGGTCATACCGCCGCAAGCCAGATCTACGGGTCGAGGGCTGCTAC 660  
Db 601 CTGCGACCGGAGGTCATACCGCCGCAAGCCAGATCTACGGGTCGAGGGCTGCTAC 660  
QY 661 GTGCTGCGCTGCTGCGCGCCGCTTTCGCGAGATGATCAAGTATGATGATACGCC 720  
Db 661 GTGCTGCGCTGCTGCTGCTGCTGCAACCCGAGATCTCGAAGTGTGATGACACGCC 720  
QY 721 GACAGAGATGTTCTTCAGAGCCGCGCGCGCTTTCGATGATTTTCGCGCCGACGCC 780  
Db 721 GACAGAGATGTTCTTCAGAGCCGCGCGCGCTTTCGATGATTTTCGCGCCGACGCC 780  
QY 781 CGCGCCCTGCGCGAGCCGCTCCGAGACCGAAGAGGAGCTGCTGCGCCGATATGAC 840  
Db 781 CGCGCCCTGCGCGAGCCGCTCCGAGACCGAAGAGGAGCTGCTGCGCCGATATGAC 840  
QY 841 CTCGCGATGATGCGCTTTCGCGAGCGCGCGCCGATCCGCGCGCACTATTCAGCGCC 900  
Db 841 CTCGCGCGATGATGCGCTTTCGCGAGCGCGCGCCGATCCGCGCGCACTATTCAGCGCC 900  
QY 901 GAGCTAACGCGCGCTGCTGCTGATGAGTTCGCGCGCCGATCCGCTTCAGCTGATGCC 960  
Db 901 GAGCTAACGCGCGCTGCTGCTGATGAGTTCGCGCGCCGATCCGCTTCAGCTGATGCC 960  
QY 961 GCATTGCAACCGCAAAAGAGACAGAGCGAGCGCGCCGCGCTGCGCTGCTGCGCG 1018  
Db 961 GCATTGCAACCGCAAAAGAGACAGAGCGAGCGCGCCGCGCTGCGCTGCTGCGCG 1018

RESULT 4  
AY487543  
LOCUS  
DEFINITION  
AY487543 1062 bp DNA linear ENV 05-Apr-2004  
Uncultured organism clone 2A20 nitrilase (BD7753) gene, complete cds.  
ACCESSION  
AY487543  
VERSION  
AY487543.1 GI:40890290  
KEYWORDS  
ENV.  
SOURCE  
uncultured organism  
ORGANISM  
uncultured organism  
REFERENCE  
1 (bases 1 to 1062)  
Robertson, D.E., Chaplin, J.A., Desantis, G., Podar, M., Madden, M.,  
Chi, E., Richardson, T., Milian, A., Miller, M., Weiner, D.P., Wong, K.,  
McQuaid, J., Farwell, B., Preston, L.A., Tan, X., Snead, M.A.,  
Keller, M., Mathur, E., Kretz, P.L., Burk, M.J., and Short, J.M.  
Exploring nitrilase sequence space for enantioselective catalysis  
Appl. Environ. Microbiol. 70 (4), 2429-2436 (2004)  
2 (bases 1 to 1062)  
Robertson, D.E., Chaplin, J.A., Desantis, G., Podar, M., Madden, M.,  
Chi, E., Richardson, T.H., Milian, A., Miller, M., Weiner, D.P., Wong, K.,  
McQuaid, J., Farwell, B., Preston, L.A., Tan, X., Keller, M., Mathur, E.,  
Kretz, P.L., Burk, M.J., and Short, J.M.  
Direct Submission  
Submitted (26-NOV-2003) Bioinformatics, Diversa Corporation, 4955  
Directors Place, San Diego, CA 92121, USA  
FEATURES  
source  
1..1062  
/organism="uncultured organism"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:155900"  
/clone="2A20"  
/environmental\_sample  
1..1062  
/locus\_tag="BD7753"  
1..1062  
/locus\_tag="BD7753"  
/EC\_number="3.5.5.7"

ORIGIN  
Query Match 46.7%; Score 486; DB 3; Length 1062;  
Best Local Similarity 67.8%; Pred. No. 4.5e-53;  
Matches 697; Conservative 0; Mismatches 325; Indels 6; Gaps 1;  
/codon\_start=1  
/evidence=experimental  
/product="nitrilase"  
/protein\_id="FAR97490.1"  
/db\_xref="GI:40890291"  
/translation="MGVAHPKRVAAVQAAAPFLDIDASVEKAVRPIDEAGAAARLI  
AFPEITPGYPMWITGAPAMATMRGVSRYFDNISLSDSPQAEKLRRAAKRMVVV  
LGLSERDGSLSLYIAQWIIGPDGTTAKRRKLKPTAERYVFGSDSHLAVHLDVGR  
LGLALCCWEHLQPLSKYAMVQNEOVHVAHMFSLYDPAFHALGAEVNNASAKIYAVE  
GSCFVIAPCATVQAMIDELCDTPEHGFHAGGGFAVIYGPDAFLAALPDPKGL  
LYVDIDLMISVAKMADIPAGHYARADVIRLLFNRRPGRVETMALPIDAETAEAPAA  
KEPKKAPNVAFPAPVQAE"

QY 17 CGAAGTATCGCGCGCGCGCGGTGTCAGCGCGCGCGCGGTTCCTCGATCTTGACCGCACAG 76  
Db 17 CGAAGTATCGAAGTATCGCGCGCGCGGTGTCAGCGCGCGCGCGGTTCCTCGATCTTGACCGCACAG 76  
QY 77 TCGAAGAAAGCGATCGCGCTGATCGAGCAGCGCGCGCAAGCAGAGCGTGCCTGATCGCAT 136  
Db 77 TCGAAGAAAGCGATCGCGCTGATCGAGCAGCGCGCGCGCGCGCGCGCGCTGATCGCAT 136  
QY 137 TCCCAAGACTTGATGATTTCCGCTATCCCTTTTGATATAGCTGCGCGCGCGCGCTTGGG 196  
Db 137 TCCCAAGACTTGATGATTTCCGCTATCCCTTTTGATATAGCTGCGCGCGCGCGCGCTTGGG 196  
QY 137 TTCGGAAGACCTGGAATACCGGTTACCCCTGTGATCTGCGTAAAGCGCGCGCGCGCTGAG 196  
Db 137 TTCGGAAGACCTGGAATACCGGTTACCCCTGTGATCTGCGTAAAGCGCGCGCGCGCGCTGAG 196  
QY 197 GCAT-----GCAGCTTCGTCAGCGCGCTATTTGAGAAATTCGCTCGTGCAGCGCGCAAGC 250  
Db 197 GCAT-----GCAGCTTCGTCAGCGCGCTATTTGAGAAATTCGCTCGTGCAGCGCGCAAGC 250  
QY 197 CTATCATGCGCGCGCTTCGCTCGCGCTATTTGACAACTCGCTCAGACTACAGACGCCGC 256  
Db 197 CTATCATGCGCGCGCTTCGCTCGCGCTATTTGACAACTCGCTCAGACTACAGACGCCGC 256  
QY 251 AGTGCAGAGCGCTGCGGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTTATTA 310  
Db 251 AGTGCAGAGCGCTGCGGATGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTTATTA 310  
QY 257 AGCGCGAAGAGCTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTCT 316  
Db 257 AGCGCGAAGAGCTCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTCT 316  
QY 311 CGGAGCGCGCGCGCGCGCGCGCTATATAGGCGCAGCGCGCTTTCGCGCGCGCGCGCGCG 370  
Db 311 CGGAGCGCGCGCGCGCGCGCGCTATATAGGCGCAGCGCGCGCTTTCGCGCGCGCGCGCGCG 370  
QY 317 CGGAGCGCGCGCGCGCGCGCGCTTATATCGCGATGATCATCGCGCGCGCGCGCGCG 376  
Db 317 CGGAGCGCGCGCGCGCGCGCGCTTATATCGCGATGATCATCGCGCGCGCGCGCGCG 376  
QY 371 TGATCCCGCGCGCGCGCGCGCGCTTACAGCTTACAGCTTACAGCGCGCGCGCGCGCG 430  
Db 371 TGATCCCGCGCGCGCGCGCGCGCTTACAGCTTACAGCTTACAGCGCGCGCGCGCGCG 430  
QY 377 CCATGCGCAAGCG 436  
Db 377 CCATGCGCAAGCG 436  
QY 431 GAGACCGCAGCGCATCTCGCGGTGACAGATACCGCGCATCGCGCGCGCGCGCGCTTGT 490  
Db 431 GAGACCGCAGCGCATCTCGCGGTGACAGATACCGCGCATCGCGCGCGCGCGCGCTTGT 490  
QY 437 GCGACCGCTCGCATCTTGGGTGACAGAGCTTATGTTGGCGCGCGCGCGCGCTGTGCT 496  
Db 437 GCGACCGCTCGCATCTTGGGTGACAGAGCTTATGTTGGCGCGCGCGCGCGCTGTGCT 496  
QY 491 GCTGGAGGACATCCAGCATTTGGAATATACGCGTATGACCGCGCGCGCGCGCGAGAGTCC 550  
Db 491 GCTGGAGGACATCCAGCATTTGGAATATACGCGTATGACCGCGCGCGCGCGCGAGAGTCC 550  
QY 497 GCTGGGAACACTGCGACGCGCGCTGTCCAAATATGCGCATGTATGCGCGCGCGCGCG 556  
Db 497 GCTGGGAACACTGCGACGCGCGCTGTCCAAATATGCGCATGTATGCGCGCGCGCGCG 556  
QY 551 ACCTGCGCTGTCGCGCGCGCGCGCTTCAAGCTTATCGCGCGCGCGCGCGCGCGCGCG 610  
Db 551 ACCTGCGCTGTCGCGCGCGCGCGCGCTTCAAGCTTATCGCGCGCGCGCGCGCGCGCGCG 610  
QY 557 ATGTCGCGCGCGCTGCG 616  
Db 557 ATGTCGCGCGCGCTGCG 616  
QY 611 AGGTCAATACCG 670  
Db 611 AGGTCAATACCG 670  
QY 617 AAGTGAACAATGCG 676  
Db 617 AAGTGAACAATGCG 676  
QY 671 CGTGCAGACCGTTTCGCGCGAGATGATCAAGTATTTGATGATACGCCGCAAGAGAGA 730  
Db 671 CGTGCAGACCGTTTCGCGCGAGATGATCAAGTATTTGATGATACGCCGCAAGAGAGA 730  
QY 677 CGTGCAGACCGTTTCGCGCGAGATGATGAGAACTTCGCGATACGCCGCAAGAGAGATC 736  
Db 677 CGTGCAGACCGTTTCGCGCGAGATGATGAGAACTTCGCGATACGCCGCAAGAGAGATC 736  
QY 731 TGTTCCTCAAGCGCGCGCGCGCGTTTTCGCGATATTTTCGCGCGCGCGCGCGCGCTG 790  
Db 731 TGTTCCTCAAGCGCGCGCGCGCGTTTTCGCGATATTTTCGCGCGCGCGCGCGCGCTG 790  
QY 737 AGTTCTGATGCGCGCGCGCGCGCTTTCGCGATATTTTCGCGCGCGCGCGCGCGCTG 796  
Db 737 AGTTCTGATGCGCGCGCGCGCGCTTTCGCGATATTTTCGCGCGCGCGCGCGCGCTG 796  
QY 791 CCGAGCGCGCTCCCGAGAGACCGAAGAGGAGCTCTGCGCGCGCGCATATCGCGCATCG 850  
Db 791 CCGAGCGCGCTCCCGAGAGACCGAAGAGGAGCTCTGCGCGCGCGCATATCGCGCATCG 850  
QY 797 CCGCGCGCGCTCGATGA 856  
Db 797 CCGCGCGCGCTCGATGA 856





DEFINITION Uncultured organism clone 2A17 nitrilase (BD5263) gene, complete  
cde-  
ACCESSION AY487522  
VERSION AY487522.1 GI:40890248  
KEYWORDS ENV.  
SOURCE uncultured organism  
ORGANISM uncultured organism  
REFERENCE 1 (bases 1 to 1077)  
AUTHORS Robertson,D.E., Chaplin,J.A., Desantis,G., Podar,M., Madden,M.,  
Chi,E., Richardson,T., Milan,A., Miller,M., Weiner,D.P., Wong,K.,  
McQuaid,J., Farwell,B., Preston,L.A., Tan,X., Sneed,M.A.,  
Keller,M., Mathur,E., Kretz,P.L., Burk,M.J. and Short,J.M.  
Exploring nitrilase sequence space for enantioselective catalysis  
Appl. Environ. Microbiol. 70 (4), 2429-2436 (2004)  
TITLE  
JOURNAL  
REFERENCE 2 (bases 1 to 1077)  
AUTHORS Robertson,D.E., Chaplin,J.A., Desantis,G., Podar,M., Madden,M.,  
Chi,E., Richardson,T.H., Milan,A., Miller,M., Weiner,D.P., Wong,K.,  
McQuaid,J., Farwell,B., Preston,L.A., Tan,X., Keller,M., Mathur,E.,  
Kretz,P.L., Burk,M.J. and Short,J.M.  
Direct Submission  
TITLE  
JOURNAL  
REFERENCE Submitted (26-NOV-2003) Bioinformatics, Diversa Corporation, 4955  
Directors Place, San Diego, CA 92121, USA  
Location/Qualifiers  
FEATURES  
source  
1. 1077  
/organism="uncultured organism"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:155900"  
/clone="2A17"  
/environmental\_sample  
1. 1077  
/locus\_tag="BD5263"  
1. 1077  
/locus\_tag="BD5263"  
/EC\_number="3.5.5.7"  
/codon\_start=1  
/evidence=experimental  
/product="nitrilase"  
/protein\_id="AA97469.1"  
/db\_xref="GI:40890249"  
/translation="MALTHPKLVAAVQAAPFLDVAADVAKVRLLIDEAANGSLV  
APPEPTPGYPPWIMIGSPAMALMRGVSVSPDNSLSYDRQAEPLRDAKRRKLTIV  
MGLSERAGSLITYQMILGPNGETVAKRKLKTFHARIVTFGSGSHLVNHLPIGR  
LGLACWEHLQPLSKYAMVAQNEIHVAAPSFSLYDPFHALGAENVNAASQIAYE  
GSCFVAPCAVISOEMIDLMCDTPDKQLTHVGGFVITGPDPGARIGDLADPQGI  
VVAIDIGMIPAKAADPAHYARPDVTRLFNHNPANRVETLVLPVDQRDIDARV  
EAAAPQARPATGNEPDAKPMALAE"  
ORIGIN  
Query Match 43.7%; Score 455.2; DB 3; Length 1077;  
Best Local Similarity 67.9%; Pred. No. 3.9e-49;  
Matches 653; Conservative 0; Mismatches 303; Indels 6; Gaps 1;  
QY 17 CGAAGTATCGGGGGGGGGGGTGGAGCGCGCGGGTTCCTCGATCGACCGCAG 76  
DB 17 CGAAATTGAAAGTGGCCGGTGGAGCGACTCCGGTTCCTCGATCGATGCCGAG 76  
QY 77 TCGAAGAAAGCATCGGCTTGATCGAGCAGCGCGCAAGCAGACGTCGCGCTGATCGCAT 136  
DB 77 TGAACAAAGCGGTGGCGGTATCGACGAGCGGAGCAAAAGGCTCCAGTCTGGTGCAT 136  
QY 137 TCCCAAGAGATTGGATTCCCGGCTATCCCTTTGGATATGGCTGGGGCGCGCTTGGG 196  
DB 137 TCCCGAGAGCTGATTCGCCGGCTATCCGTTTGGATCTGGCTTGGCTCGCGCGCTGGG 196  
QY 197 GCAT-----GCGCTTCGTCAGAGCGATATTTCGAAATTCGCTCGCGCGGCAAGC 250  
DB 197 CATCATGCGCGGGTGTGTCTCTCGATTTCGATTAATTCGCTCAGTATGACAGCGCGC 256  
QY 251 AGTGGACAGCGCTGGCGGATCGGCGCGCGCAAGCATGATGTCGCGCGCTATA 310  
DB 257 AGGCAGAGCGCTGGCGGAGCGCGCAAGCGCAAACTGACCGTGTATGAGGCGCTGT 316

QY 311 GCGAGCGCGCGGGCGGAGCGCTCTATATGGCGCAGCGCATCTTGGCCCGGATGGCAGTC 370  
DB 317 CCGAGCGCGCGCGGGTGGAGCTTTAATCGCGAGTGGATCATTTGCCAATGGGAGCA 376  
QY 371 TGAATCGCGCGCGCGCGCAAGCTCAAGCTTACCCATCGAGCGCACCGTGTTCGGGAGG 430  
DB 377 CCGTCCGACAGCGGGCGCAAGCTCAAGCCCAACCCATCGAGCGCACCGTGTTCGGGAGG 436  
QY 431 GAGACGCGACCATCTTCGGGGTGCACGATACCGCCATCGGGCGCTTCGCGCTCTGT 490  
DB 437 GTACCGCAGCGCCACCTGGCGGGTACACAACTTTCACATCCGACCGCTCGGCTGTGCT 496  
QY 491 GCTGGAGGACATCAACGACCATTTGCAAAATAGACGATAGACGCGCGCAAGCAAGATCC 550  
DB 497 GCTGGAGGACATCTCAAGCGCTCTCCAAATAGCGATGACGCCCAAGCAAGATCC 556  
QY 551 ACCTGCGCTGTGGCGGAGCTTCAAGCTTATCGCGGACATGCGCTTATGCGCTGACCGG 610  
DB 557 ACCTGCGCGGATGCGCGCTCTCTCGCTTACGACCGGTTGGCGACCGCTCGCGCGC 616  
QY 611 AGGTCAATACCGCGCGCAAGCATATACCGGTCGAGCGCGCTACGTCGTCGCT 670  
DB 617 AAGTCAACACGAGCGAGCGCATATACGCGGTAAAGTTCCTCTTGTGTCGCGC 676  
QY 671 CGTGGCGACCGCTTTCGCGGAGATGATCAAGGTATTGGATGATACCGCGCAAGAGAGA 730  
DB 677 CATGTGCGGATCTCGCAGGAAATGATGATTTATGTCGATACCCCGCAACATC 736  
QY 731 TGTTCCTCAAGCGCGCGCGGCTTTTGCATGATTTTCGGGCGCGACCGCGCGCTGG 790  
DB 737 AGCTTATTCACGTCGCGTGGCGGCTTACCGGTATCATGAGCCGAGCGGTGCGCGATCG 796  
QY 791 CCGAGCGCTCCCGGAGACCGAAGAGAGACGTCGTGTCGCGCATTCGATCTCGCGATGA 850  
DB 797 GCGACAAAGCTCGCGCGCAGATCAGAAAGCATTTCTATGCCATCGATCTCGCATGA 856  
QY 851 TCGCGTTGGCAAGCGCGGCGCGGATCGCGCGGCACTATTTCACGCGCGCATGACG 910  
DB 857 TCCGATTCGCAAAAGCTGGCGCGCATCTGCGGCGCATATGCGCGACCGCATGATCC 916  
QY 911 GCGTGTGCTGATCGACGTCGCGCGCAAGCGTGTACGCTGATGTCGCGATGGAAC 970  
DB 917 GCTTCTGTTCACAATCGTCCGCAATGGGTGAAACCTCTGTCCTCCGTTGATC 976  
QY 971 CG 972  
DB 977 AG 978  
RESULT 7  
AY487536 1017 bp DNA linear ENV 05-APR-2004  
LOCUS  
DEFINITION Uncultured organism clone 2A29 nitrilase (BD5279) gene, complete  
cde-  
ACCESSION AY487536  
VERSION AY487536.1 GI:40890276  
KEYWORDS ENV.  
SOURCE uncultured organism  
ORGANISM uncultured organism  
REFERENCE 1 (bases 1 to 1017)  
AUTHORS Robertson,D.E., Chaplin,J.A., Desantis,G., Podar,M., Madden,M.,  
Chi,E., Richardson,T., Milan,A., Miller,M., Weiner,D.P., Wong,K.,  
McQuaid,J., Farwell,B., Preston,L.A., Tan,X., Sneed,M.A.,  
Keller,M., Mathur,E., Kretz,P.L., Burk,M.J. and Short,J.M.  
Exploring nitrilase sequence space for enantioselective catalysis  
Appl. Environ. Microbiol. 70 (4), 2429-2436 (2004)  
TITLE  
JOURNAL  
REFERENCE 2 (bases 1 to 1017)  
AUTHORS Robertson,D.E., Chaplin,J.A., Desantis,G., Podar,M., Madden,M.,  
Chi,E., Richardson,T.H., Milan,A., Miller,M., Weiner,D.P., Wong,K.,  
McQuaid,J., Farwell,B., Preston,L.A., Tan,X., Keller,M., Mathur,E.,  
Kretz,P.L., Burk,M.J. and Short,J.M.  
Direct Submission

**JOURNAL** Submitted (26-NOV-2003) Bioinformatics, Divera Corporation, 4955  
Directors Place, San Diego, CA 92121, USA  
**FEATURES** Location/Qualifiers

FEATURES	Location/Qualifiers
source	1. .1017

```

/environmental_sample
1. .1017
gene

```

1. .1017 /locus\_tag="BD5279"  
CDS

```

/ locus_tag="BD5279"
/ EC_number="3.5.5.7"
/codon_start=1
/evidence=experimental
/product="nitritease"
/protein_id="AA097483.1"
/db_xref="cgi:10890277"
/translation="MGIEHKRGVAVVOAFAVNPALDRLRSIDAEILIEEAAANGARLTA
APPEVFLPGYPMWHLITGDSPPALMAGIGGVORFENSLAYPSIDARLRAAVKARLTA
IGLSERGGSGLYIMQIVGDDGELIAKRRLRPTHAERIVYEEGGSDIAVHREDGIG
RLGAGCMEHSLQIKSTKAAVNOEDVYHVAWSPFSIYDFEALDAETVNNASVRVNA
EGGSFVLAFCPTGSOQAMIDELCRPDKHALIHAGGSPRALITYPDSSSLAEKLAPOGSE
LALVLDIDIGALGVAKNADPAHGISRPDVTLLLNKKPYKRVHEALPGDVTAPADVT
"
/aaas

```

## ORIGIN

Query Match	43.7%	Score 454.6	DB 3	Length 1017
Best Local Similarity	67.5%	Pred. NO. 4.7e-49		
Matches 672; Conservative	0	Mismatches 314	Indels 9	Gaps 2

QY	17	CGAAGTATCGCGGCGCGGCGGTGACAGGCGCGCGCGGTTCCTCGANTCTGCACGACAG	76
Db	17	CAAAATACCGCGTTGCCGCGCGGTGACAGGCTGGCGCGGCTGGCTGCACCTCGANTCGCTCGA	76
QY	77	TCGAGAAAGCATTCGGGCTGATTCGAGCAGGGGGGCCAAGCAGAGCTGGGCTGATTCGAT	136
Db	77	TCGCAAGAGGCTATTGCACTGATTCAGAGGAGGGCGCGCAACGGGGCCAGATTGATTCGAT	136
QY	137	TCCGAGAGACTTGGATTCCCGGCTATCCCTTTTGGATTATGGCTGGGCGCGCGGCTTGG-	195
Db	137	TCCCGGAGGTTTTCATCCCGGGCTACCCCTGGACATTTGGCTGCACCTCGCGGCTGGG	196
QY	196	-----GGCANTCGCTTTCGTCACGCGCTATTTTCGAGATTTCGCTGGTCGCGGACAGAAC	250
Db	197	CGATCGGCGCGGCTTCGTGACGAGGTATTTCGCAATTCCCTCGCTTATGATGTGCGC	256
QY	251	AGTGGCAGGGCCCTGGCGGATGGCGCGCGCGCCGACAGGATGCATGTCGTGGCCGGCTATA	310
Db	257	AGGCGGAGCGGCTCCGCGCAGCGGTGCGCAAGCGCGCTACCGCGGTGATCGGCTTT	316
QY	311	GCGAGCGCGCGGGCGCGAGCCTCTATATAGGAGCAGGCGCATTTGCGGCCCGCATGGCGATC	370
Db	317	CGAGCGGAGCGGGCGGAGCCTTACATCGGCAATGGCTGTGGGCCCGACGGCGAGA	376
QY	371	TGATCGCGCGCGCGCGCAAGCTTCAAGCTTACCATGCGGAGCGACCGTGTTCGGCGAG	430
Db	377	CGATTCGGGAACGCGCGCAAGCTTCGTCGCAAGCATGCGAGCGACCGGTCTATGGCCAGG	436
QY	431	GAGACGCGAGCATCTCGCGGTGACAGA---TACCGGCATCGGGGCGCTCGCGCGCTCT	487
Db	437	GCGACGCGAGGATCTGGCGGCTCCATATACCGGGCCGGAATTCGAGAGGCTTCGGCGCGCT	496
QY	488	GTTCGCTGGAGACATCCAGCCATTGTGAAATACGCCATGTATGCGCGCCGACGAAACAGG	547
Db	497	GCTCTGGGAACATCTGCAACCGTGTGCGAAATATGCGATGTATAGCCCAAGACAGGAGG	556
QY	548	TCCAGTCGCGGTCGTTGCGGCGAGCTTCAAGCTCTATTGCGGGATGAGGCTTATGCGGCGAC	607
Db	557	TTCATGTGGGCTATGAGCGCGAGTTTTCGTGCTCTACAGATCCCTTTGCCCGCGCTTCGGCG	616
QY	608	CGAGGTCATATACCGCGCGAAGCCAGATCTTACGCGGTCGAGGAGCGGCTGTTACTGCTGG	667

Db	617	CGAGGCTGGAACAATGCGCGCTCTCCGGGTCTATGCGGTGGAAGGCTCTGCTGTGTGGTGG	616
Qy	668	CGTCGTGCGGCAACCGTTTCCGCCGAGATGATCAAGTATTGTTGATGATACGCCGCAAG	727
Db	677	CGCGTCCGCGGACCGTCTCCAGGCCATGATTCGATGCTGTGTGCAACGCCGCCACAAAG	736
Qy	728	AGATGTTCTCAAGAGCGCGGCGGGTTTTCGATGATTTTCGGGCCCGACGCGCGCGCC	787
Db	737	ACGGCGTCTTCATGCGCGCGGTGCTTTGTCGCGATCTACGGGCCCGACGCGCATTTGCTG	796
Qy	788	TGGCCGAGCCGCTCCCGAGACCGAAGAGGGACTGCTGATCGCCGATATCGACTTCGCA	847
Db	797	TGGCCGAAGAAGCTGCGCCCGAACCGAGAGGGCTGTGCTTTACGCCGACATGATCTCGGCG	856
Qy	848	TGATTCGGTGTGCCAAGCGCGCGCGCATTCGCGCGGCCCATTTTCAAGCGCCGCAAGTAA	907
Db	857	CGATCGGCGTGCAGAAACGCGCGCCACCCCGAGGGGCAATTATTCGCGCGCGATGTCA	916
Qy	908	CGCGGCTGCTGTGATCGACGTTCGCGCCCAACGCGTCGTACAGCTTGATGCGCGATTCG	967
Db	917	CGGGGCTGTGTGTGAACAACAGCCCTTACAAGAGCGGTGAAGCATTTTGTCTTTGCCCGCGG	976
Qy	968	AACCGCAAAACGAAGACAAAGGCGACGCGCCCGG	1002
Db	977	ATACCGTGGCGCCTGCGCGATGTGATTCGCGCGG	1011

RESULT	8
LOCUS	AY487558
DEFINITION	AY487558 1017 bp DNA linear ENV 05-APR-2004
ACCESSION	AY487558
VERSION	AY487558.1
KEYWORDS	Gt:40890320
SOURCE	ENV.
ORGANISM	uncultured organism
REFERENCE	uncultured organism
AUTHORS	unclassified; environmental samples. 1 (bases 1 to 1017) Robertson,D.E., Chaplin,J.A., Desantis,G., Podar,M., Madden,M., Chi,E., Richardson,T., Milan,A., Miller,M., Weiner,D.P., Wong,K., McQuaid,J., Farrell,B., Presoon,L.A., Tan,X., Sned,M.A., Keller,M., Maturr,E., Kretz,P.L., Burk,M.J. and Short,J.M. Exploring nitrilase sequence space for enantioselective catalysis Appl. Environ. Microbiol. 70 (4), 2429-2436 (2004)
TITLE	2. (bases 1 to 1017) Robertson,D.E., Chaplin,J.A., Desantis,G., Podar,M., Madden,M., Chi,E., Richardson,T.H., Milan,A., Miller,M., Weiner,D.P., Wong,K., McQuaid,J., Farrell,B., Presoon,L.A., Tan,X., Keller,M., Maturr,E., Kretz,P.L., Burk,M.J. and Short,J.M. Direct Submission
JOURNAL	Submitted (26-NOV-2003) Bioinformatics, Diversa Corporation, 4955
REFERENCE	Directors Place, San Diego, CA 92121, USA
AUTHORS	location/Qualifiers
TITLE	1..1017
JOURNAL	/organism="uncultured organism"
FEATURES	/mol_type="genomic DNA" /db_xref="taxon:155900" /clone="ZA31" /environmental_sample 1..1017 /locus_tag="BD5287" 1..1017 /locus_tag="BD5287" /EC_number="3.5.5.7" /codon_start=1 /evidence=experimental /product="nitrilase" /protein_id="FAR97505_1" /db_xref="Gt:40890321" /translation:"WGIGHKRVAVVOAAPMLDLBASYSKIALTEPAAGAKIIL APEAFIPGWPYIMLVGRGFVORYFNSTSYDSPOAERRLNAVKKAGMTAV



Db 137 TCCCGAAGCCTTCATTCGCCGGCTACCTTGCGACATCTGGCTGCACTCACCGGCTGGG 196  
Qy 197 GCATGCCG-----TTGCTCAGAGGCTATTTCGGAATTCGCTGCTGGCGGCGAGCAAGC 250  
Db 197 CGATCGCGCGGCTTCTGTGAGGCGCTATTTCGCAATTCCTCGCCCTACACAGAGCCAC 256  
Qy 251 AGTGGCAGAGCCCTTGCGGAGTGGCGGCGCGCGCAACGCGATGATGCTGGCGGCGATATA 310  
Db 257 AGCGGGAAGAGCTCGACAGAGCCTGCGGGAAGGCGCAAGCTACCGGCTGATTCGCTCT 316  
Qy 311 GCGAGCGCGCGGCGCGAGCCTTATATATGCGCGCAAGCGATCTTCGCGCCCGATGCGATC 370  
Db 317 CCGAGCGCGAGCGGCGGAGCCTTATCTCGCGAGTGGCTGATCGGCGCCGACGCGTGA 376  
Qy 371 TGATCGCGCGCGCGCGCAAGCTCAAGCTTACCATGCGAGAGCGACCGTGTTCGCGAG 430  
Db 377 CCATCGCAAAAGCGCGCGCAAGCTGCGCGCGCAACCTACGCGACCGACCTTATGCGAG 436  
Qy 431 GAGACGGAGCGCATCTGCGCGGTGACGA---TACCGGACGCGGCGCTCGCGCGCTCT 487  
Db 437 GCGAGCGAGCGATCTCGCGCTCAATGAGCGGCGCGAGATCGCGCTCGCGCGCTGT 496  
Qy 488 GTTCTGGAAGCACAATCCAGCATTTGTGAATAAGCCATGTAAGCGCGCGCGACGAGAG 547  
Db 497 GCTGCTGGAGCATCTGAGCGCGCTGTGGAATTCGCAATGTAAGCGCGCGAGAGAG 556  
Qy 548 TCCAGCTGCGCTCGTGGCGAGCTTCAAGCTTATCGCGGATGAGCTTATGCGCTCGAG 607  
Db 557 TACATGTCGCGCGCTGCGCGAGCTTCTGCTCAAGATCCCTTGGCGCTGCGCTGGGCG 616  
Qy 608 CCGAGGTCAATACCGCGCGCAAGCAGATCTPACGCGGTGAGAGGCGCGCTGTAAGTCTG 667  
Db 617 CCGAGGTGAACAAGCGCGCTCGCGCATTAAGCGGGAAGGCTCTGCTTGTGCTCG 676  
Qy 668 CGTGTGCGCGACCGCTTTCGCGGAGATGATCAAGTATTGATGATACGCGCGCAAG 727  
Db 677 CACGCTGCGGAGCGGTCTCGAGGCGATGATGAGAGCTTGGCATGCGCGCGAGCAAGC 736  
Qy 728 AGATGTTCTTCMAAGCGCGCGCGGCTTTCGCAATGATTTTCGAGCGCGCGCGCGCC 787  
Db 737 ACGGCTGCTGATGCGCGCGCGCGGCTTTCGCGATCTACGCGCGCGAGCGAGCA 796  
Qy 788 TGGCGGAGCGCGCTCCCGGAGACCGAAGAGAGGATGCTGCTGCGCGATATGCACTCG 847  
Db 797 TCGGCGCAAGCTCTCCCGCGAGAGGCGCTGCTGATGCGGAGATGATCTGAGCG 856  
Qy 848 TGATGCGGTTGGCGAAGCGCGCGCGATCGCGCGCGCACTATTTCAGCGCGCGAGTAA 907  
Db 857 CGATCGCGCTGCGCAAGAGCGCGCGATCCCGCGGCACTTTCGCGCGCGAGCTGCA 916  
Qy 908 CGCGGCTGCTGCTGATGATGATCGTCCGCGCAAGCGCTGTCAGCGCTTGAATGCG 967  
Db 917 CGCGGCTGCTGCTCAACGAGAAAGCGCAAGCGCGTGAAGATGTCGCGTCCGCTG 976  
Qy 968 AACCGCAAAAGAGAGAGAGGCGAGCGCGCGCC 1001  
Db 977 ACACGTCGAGCGCGCTGCGAGCTGCGCGGCGAGC 1010

RESULT 10  
LOCUS AY487559 1005 bp DNA linear ENV 05-APR-2004  
DEFINITION Uncultured organism clone ZA36 nitrilase (BD7051) gene, complete cds.  
ACCESSION AY487559  
VERSION AY487559.1 GI:40890322  
KEYWORDS ENV.  
SOURCE uncultured organism  
ORGANISM uncultured organism  
REFERENCE 1 (bases 1 to 1005)  
AUTHORS Robertson, D.B., Chaplin, J.A., Desantis, G., Podar, M., Madden, M., Chl, E., Richardson, T., Milan, A., Miller, M., Weiner, D.P., Wong, K.,

McQuaid, J., Farwell, B., Preston, L.A., Tan, X., Sneed, M.A., Keller, M., Mathur, E., Kretz, P.L., Burk, M.J. and Short, J.M.  
Exploring nitrilase sequence space for enantioselective catalysis Appl. Environ. Microbiol. 70 (4), 2429-2436 (2004)  
2 (bases 1 to 1005)  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
SUBMITTED (26-Nov-2003) Bioinformatics, Diversa Corporation, 4955 Directors Place, San Diego, CA 92121, USA  
LOCATION/Qualifiers  
FEATURES  
source  
1..1005  
/organism="uncultured organism"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:155900"  
/clone="2A36"  
/environmental\_sample  
1..1005  
/locus\_tag="BD7051"  
1..1005  
/locus\_tag="BD7051"  
/BC\_number="3.5.5.7"  
/codon\_start=1  
/evidence=experimental  
/product="nitrilase"  
/protein\_id="AAR97506.1"  
/db\_xref="GI:40890323"  
/translation="MGTKHPAFKAAYVOAPEWLDLDRIVDTXLTALIEBAAGAKLI  
APETWPGVPMHIVGTGPAWLSRGVQYFPNSLYADSPQARIDAKKXITV  
LGISEBEGSLYISQWLIQDGTIAKRLTPTHTYRRTYFGDGSHTIAVHRADIG  
RLALCMWHTIOPUTRIKAMTQNEQVHVAWPSFMTIEPFAHLMGVNNAASLTIV  
EGSCFVLGACAVISQAMVDEMDTEDEKRALVHAGGHAIVFGDGLAKIPIETQG  
LTVADIDLAGIIVAKNAADPAHYSPDVTRLFNNKPARVEYFSLPVDVETPPQ  
"  
ORIGIN  
Query Match 43.2%; Score 450; DB 3; Length 1005;  
Best Local Similarity 69.4%; Pred. No. 1.8e-48;  
Matches 644; Conservative 0; Mismatches 275; Indels 9; Gaps 2;  
Qy 28 GGGCGCGGCTGAGAGCGCGCGCGGCTTCTGATGATCGACCGCAAGTCGAGAAAGCG 87  
Db 28 GCGCAAGTGTTCAGAGCGCGCGCGGAATGCTGATTCGACCGACGTCGACAAAGC 87  
Qy 88 ATCGGCTGTGAGAGCG 147  
Db 88 ATCGGCTGTGAGAGCG 147  
Qy 148 TGATTCGCGGCTATCCCTTTGATATGCTGCGCGCGCGCGCGCGCGCTTGGG-----GCATG 201  
Db 148 TGATTCGCGGCTATCCGCTGAGCAATCTGGGTGCGCACGCGCGGTGGGATCAGCGC 207  
Qy 202 CGCTTGTGAGCGCTATTTCGGAATTCGCTGCTGCGCGCGCGCAAGTGGCAAGCGC 261  
Db 208 GGGCTTGTGAGCGCTATCTGCAATTCACGCGCTTACGACAGCGCGCGCGCGCGC 267  
Qy 262 CTGGCGATGCG 321  
Db 268 ATCGCGAGCG 327  
Qy 328 GGGCGAGCTTATATGAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 381  
Db 328 GGTGAGCGCTTATATCTGCGAGTGGCTGATTTGGGCGCGCGCGCGCGCGCGCGCG 387  
Qy 382 CGCGCAAGCTCAAGCTTACCATGCGAGAGCGAGCGCGCGCGCGCGCGCGCGCGCGCG 441  
Db 388 CGCGCAAGCTGCG 447  
Qy 442 CATCTGCGGCTGACGA---TACCGCATGCGCGCGCTCGCGCGCGCTTCTTCTGGAG 498  
Db 448 CATCTGCGGCTGACGAAGCTGCTGACATGCGCGCGCGCGCGCGCGCTCTCTCTGGAG 507

QY 499 CACATCCAGCCATTGTGCAATATCCAGTATGCCCGCCGACGAACAGTCCACGTCCG 558  
Db 508 CACATCCAGCCCGCTGACCAATATCGCATGTATGCCGGAACGAGAGTCCACGTCCG 567  
QY 559 TCGGCGCGAGCTTCAGCTCTATCGCGGCATGCGCTATGCGCTTCGGAACGGAGTCAAT 618  
Db 568 GCTGCGCGAGCTTCAGCTCTATCGAGCGCTTCGCCCGCGCTCGCTGGGAAGTCAAC 627  
QY 619 ACCGCCCAAGCCAGATCTACGCGGTGAGGCGCGCTCTACGTCGTGCGCTGCGCG 678  
Db 628 AATGCGCGAGCAAGACTTACGCGGTGAGGCGCTCGCTTCGTCGTGCGCGCATGCGCG 687  
QY 679 ACCGTTTCGCGCGAGTATGATCAAGTATGATGATACGCCCGCAAGAGATGTCCTC 738  
Db 688 GTGATCTCGCAGGAGTGTGACGAATGTGCAACACCGAGCAAGCGCGCGCTGCTC 747  
QY 739 AAGCGCGCGCGCTTTTTCATGATTTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 798  
Db 748 CATGCG 807  
QY 799 CTCCCGAGACCGAAGAGGAGTCTGCTGCTGCGATTCGACCTTCGCGATGATCGCTTG 858  
Db 808 ATTCCGAGACCCGAGGAGGCTCTATGCGCATGACCTTCGCGCGCATTTGCGCGTG 867  
QY 859 GCCAAGCG 918  
Db 868 GCCAAGACCG 927  
QY 919 CTGGATGACGTCTCGGCGCGCAACGCGCTG 946  
Db 928 TTCAACAACAGCGCGCGCGCGCGCGCGTG 955

RESULT 11  
AY487430 1047 bp DNA linear ENV 05-APR-2004  
LOCUS AY487430  
DEFINITION Uncultured organism clone 2A15 nitrilase (BD5338) gene, complete cds.  
ACCESSION AY487430  
VERSION AY487430.1 GI:40890064  
KEYWORDS ENV.  
SOURCE uncultured organism  
ORGANISM uncultured organism  
REFERENCE 1 (bases 1 to 1047)  
AUTHORS Roberton,D.E., Chaplin,J.A., Desantis,G., Podar,M., Madden,M.,  
Chl,E., Richardson,T., Milan,A., Miller,M., Weiner,D.P., Wong,K.,  
McQuaid,J., Farrell,B., Preston,L.A., Tan,X., Sneed,M.A.,  
Keller,M., Machur,E., Kretz,P.L., Burk,M.V., and Short,J.M.  
Exploring nitrilase sequence space for enantioselective catalysis  
Appl. Environ. Microbiol. 70 (4), 2429-2436 (2004)  
2 (bases 1 to 1047)  
AUTHORS Roberton,D.E., Chaplin,J.A., Desantis,G., Podar,M., Madden,M.,  
Chl,E., Richardson,T.H., Milan,A., Miller,M., Weiner,D.P., Wong,K.,  
McQuaid,J., Farrell,B., Preston,L.A., Tan,X., Keller,M., Machur,E.,  
Kretz,P.L., Burk,M.V., and Short,J.M.  
Direct Submission  
Submitted (26-NOV-2003) Bioinformatics, Diversa Corporation, 4955  
Directors Place, San Diego, CA 92121, USA  
FEATURES  
source  
1..1047  
/organism="uncultured organism"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:155900"  
/clone="2A15"  
/environmental\_sample  
1..1047  
/locus\_tag="BD5338"  
1..1047  
/locus\_tag="BD5338"  
/EC\_number="3.5.5.7"  
/codon\_start=1

/evidence=experimental  
/product="nitrilase"  
/protein\_id="AAR97377.1"  
/db\_xref="GI:40890065"  
/translation="MPSKQFVAAYAOAPVFLDEGATSKGSLIEASNGAKLIA  
FPEFTWIPGRHWIWLDSFANGMFPVRFEDNSLMLSEBQAKRNQALANKITVWGY  
SERSGSLYWGSIINDKGETITRRLKXTHYERLVFEGDSDSHLCVMTLETGRVGA  
MCCWEHLQPLSKYAMYSDQIHLIASPFSLYRGAAYALGPBLNNAASQMYAEQC  
FVLAPCATYSKEMIEMIDPREKPLLBEGGFTWYIGDPGRPLARPLPENEGLLYA  
DILGIMISMAKAADPAHYARPDVTRLFNSAPNARVEYINPASPTESLKDGMKQ  
MEARQKALIREMI"

## ORIGIN

Query Match 43.1%; Score 448.6; DB 3; Length 1047;  
Best Local Similarity 67.7%; Pred. No. 2,7e-48;  
Matches 628; Conservative 0; Mismatches 299; Indels 0; Gaps 0;

QY 20 AGTATCG 79  
Db 17 AATTTAGATGCTCGTCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 76  
QY 80 AGAAGCGATCGCGCTGATGAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 139  
Db 77 GCAAGGCACTCCCTCATAGAGAGCGCGCTTCAATGAGCGCAAGCTATTGCTTCC 136  
QY 140 CAGAGACTTGATTCGCCGCTATCCCTTTTGATATGCGTGGCGCGCGCGCGCGCGCG 139  
Db 137 CGGAACGTGTGATTCGCGCTACCCCTGATGATCTGCTGATGACTCAACCGCTTGGGCA 136  
QY 200 TGGCTTCGTCAAGCGCTATTTGAGAAATTCGCTGTCGCGCGCGCGCGCGCGCGCG 259  
Db 197 TGGCGTTGTCAGCGCGCTATTTTGAACAATCGCTCATGCTGGGTATGAGCGCAAGC 256  
QY 260 CCTTGGCGATGCG 319  
Db 257 GCATGAACCAAGCTGCG 316  
QY 320 CGGCGCGAGCGCTCTATGATGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 379  
Db 317 GTGGCGCGAGCGCTCTATGATGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 376  
QY 380 CGCGCGCGAAGCTCAAGCGCTTACCGATGCGAGCGCGCGCGCGCGCGCGCGCGCG 439  
Db 377 CGCGCGCGAAGCTCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 436  
QY 440 GCCATCTCGCGCGTGCAGATACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 499  
Db 437 GCCATCTTTCGATATGATACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 486  
QY 500 ACATCCAGCCATTGTGAAATATGCGCATGTACCGCGCGCGCGCGCGCGCGCGCGCG 559  
Db 497 ATTGAGCGCGCTCAGCAAAATATGCAATGATCTCAGGATGAACAATTCATTTGCT 556  
QY 560 CGTGGCGAGCTTCAGCTCTATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 619  
Db 557 CTTGGCGAGCTTTTGTATATCGGAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 616  
QY 620 CGCGCGCAAGCCAGATCTACGCGGTGAGGCGCGCGCGCGCGCGCGCGCGCGCGCG 679  
Db 617 AGCGCGCGCGCAAAATGATGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 676  
QY 680 CGGTTTCGCGCGAGTATCAAGTATTTGTTGATACGCCCGCAAGAGATTTCTCTCA 739  
Db 677 CGCTTCGAAGGAGATGATGAATCTGATATATATATCCCAAGAAAGCGCGCTCTGC 736  
QY 740 AGCGCGCGCGCGGTTTTCATGATTTTTCGCGCGCGCGCGCGCGCGCGCGCGCGCG 739  
Db 737 TGAAGGTGCGCGGCTTCAACATATTTAAGCGCGCGCGCGCGCGCGCGCGCGCGCG 736  
QY 800 TCCCGAGACCGAAGGAGCTGCTGCTGCGCGATATGCACTTCGCGCATGATCGCTTGG 859  
Db 797 TGCTGAAAAACGAGGAGGCTGCTATATGCGCGATTTGACCTGGGACATGATTTCAATGG 856

Qy	860	CCAAGCGCGGAGCCCATCCGGCGGGGCACTATTACGGCCGCACTAAGCGGCTGCTGC	919
Db	857	CCAAAGCTGCGCGCGGACCCGGGAGGTCACTAAGCAGCGCCGGATGTCACTGGCTCACTAT	916
Qy	920	TGGATTCACATCTCCGGGCGCCAAAGCGCTGC	946
Db	917	TCAATTCCGCGCGCGCCAAATGCGCGTGC	943

RESULT	LOCUS	DEFINITION
12 AY487552	1062 bp	DNA linear ENV 05-APR-2004
	Uncultured organism clone 2A18	nitriliase (BD5217) gene, complete

ACCESSION	AY487552	GI:40890308
VERSION	AY487552.1	
KEYWORDS	ENV.	
SOURCE	uncultured organism	
ORGANISM	uncultured organism	

REFERENCE	AUTHORS
1 (bases 1 to 1062)	Robertson, D.E., Chaplin, J.A., Desant

**TITLE** Exploring nitrilase sequence space for enantioselective catalysts  
**JOURNAL** Appl. Environ. Microbiol. 70 (4), 2429-2436 (2004)  
**REFERENCE** 2 (bases 1 to 1062)  
**AUTHORS** Robertson, D.E., Chaplin, J.A., Desantis, G., Podar, M., Madden, M.,

**TITLE** Direct Submission  
**JOURNAL** Submitted (26-NOV-2003) Bioinformatics, Diversa Corporation, 4955  
Directors Place, San Diego, CA 92121, USA

```

FEATURES
source      Location/Qualifiers
1. .1062

```

```

organism="uncultured organism"
/mol_type="genomic DNA"
/db_xref="taxon:155900"
/clone="ZA18"
/environmental_sample
1..1062
gene
1..1062
locus_tag="BD5217"
1..1062
CDS

```

```

/locus_tag="BD5217"
/EC_number="3.5.5.7"
/codon_start=1
/evidence=experimental
/product="nitrilase"
/protein_id="AAR97499.1"
/db_xref="GI:40890309"
/translation="MSQIHPKLTVAAYVQAPAFLLDASIEKTIIRYVDEAAAGAKL
IAFPEWTIPGYEPMWITGLAPALIMRGFSRFDNSLQSGSEAEELNPAARNTIYLA
LGSEERDGGSLYIAOMITIGPGSEYVAQRRKLPHTAETVTEGGDSHLAHDLDIEV
LGAICSEWHQLPLSKYATQNEQIHVAWPSFSLYDFPALGAEVNNASAKIYAGE
GSCGVIAIPCATVSOWMIDELCDTPEKHQFLVYGGGFAYTYPDGSPLAKPLPPODEGI
LYADIDGMSIVAKRAADBPAGHYARPDVTRLFFNNRPNRRYKALPVDQAEVDSPL
KAPASPEVTILKPSQAAE"

```

Query Match	43.0%	Score 447.8;	DB 3;	Length 1062;
Best Local Similarity	67.4%;	Pred. No. 3.5e-48;		
Matches 648; Conservative	0;	Mismatches 307;	Indels 6;	Gaps 1;

QY	28	GGCGCGGCGGTGACAGGCCCGCGCGGATGTTCTCTGATCTCGACCCGACAGTCGAGAAAACG	87
Db	28	GTCGGGCGCTGACAGGCTGCCCGCCCTTTCTCGATCTCGATGCATTCATCGAAAAGACA	87
QY	88	ATCGCCCTGATCTGAGCAGCGGCGCCCAAGAGACGCGCCCTGATGCAATTCCTCCAGAGACT	147
Db	88	ATAGCGTGTGTGACGAAGCGGCTGCGGCGGGGGAAGTATATGCTTTCCGGAATCC	147

Qy	148	TGAATTC	CCCGGCTATCCCTTTTGGATATGAGCTGGGCGCGCGCTTGGGGCAT	-----G	201
Db	148	TGATATCCCGCTAC	CCATCCATGGTGGATCTGGCTGGCGCTCCGGCTGGGCATATGCGT		207
Qy	202	CGCTTCGTCCAGCG	CTATTTTCGAGATTTCCGCTCGTGGCGGCGAGCAAGCACTGGCAGGCC		261
Db	208	GGCTTCCTCTTCG	CCCTATTTTCACACATCGCTGCAATATGCGCAATCCGGAACTGAAACGG		267
Qy	262	CTGGCGGATGCGG	CGCGCGCGCACGGCATGATCTCGTGGCGGGCTATATGACGAGCGCG		321
Db	268	CTGGCGGAGCG	CGCGCGCGCACAGATATATCATCGCCCTCGGCTCTGCGAGCGCAC		327
Qy	322	GGCGGCAAGCT	CTATATATGAGGCGACGAGCATCTTCGGCCCGATGGCGATCTTGATCGCGG		381
Db	328	GGGGGCAAGT	CTTTATATCGCGCAATGATATCATCGGCGCTGGCGCGCAAA	CGTTGCAAA	387
Qy	382	CGCGCGAAGCT	CTAACGCTTACCCATACGCGGAGCGCACCGCTTCGCGCGAGGAGAGCGCGAC		441
Db	388	CGCGCGAAGCT	CTCAAGGCCCGACCGCGGAGCGCATGTATTCGCGGAAAGGCGAGTTC		447
Qy	442	CATCTCGCGGT	GCAGATACCGCCATCGGGCGCTCGCGCGCTGTGTTCGGAGCAC		501
Db	448	CATCTCGCGGT	GCAGATATCTGATATTTGAAAGATTTGGGCGCGCTTGTGTGGGAACAT		507
Qy	502	ATCCAGCCAT	TGTGTGAATATACGATGTATGCGCGCGCGACGAAACAGTTCACATGTGGCGT		561
Db	508	CTGACACCGT	TGTGTGAATATGCAATATACGCCCGAAGCGCAAAATTCACATGTGGCGCGC		567
Qy	562	TGGCGGAGCT	TTCAGCCTCTATTCGGGGATGGCGCTATGCGCTCGGACCGGAGTCAATATAC		621
Db	568	TGGCGGAGCT	TTCGCTATATACATCCCTTTGCAACGCACTCGGCGCGAGTCAATATAC		627
Qy	622	GCCGCAAGC	AGATCTTACCGCGGTGAGGGCGGCTGTACTCGTGGCGTCTGGCGGAC		681
Db	628	GCTCGAGCA	AGATCTTACCGGTGAGGGGATTCGTTCGCTATGTGGCGCGTGGCGAACG		687
Qy	682	GTTTTCGCGG	AGATATCAAGGTATTTGGTGGATACGGCCGACAGAGGATGTTCTCTCAAG		741
Db	688	GTTCGCAAG	GTGATGATGAGCTCTGCGATATCCCCGAAAAGCATCAATTCCTTTAC		747
Qy	742	GCCGCGCGG	CTTTTGCATGATTTTTCGGGCGCGACGCGCGGCTTGGCGAGCGGCTC		801
Db	748	GTGGCGGCG	GCTTCGCGCTCATTTTACGGATCCGACGGCTCGCACATGGCGCAAACTCTT		807
Qy	802	CCGGAAC	CCGAAGGGAATGCTGTGGTGGCGGATATCGACTCTGGCATGATGCGCTTGGCC		861
Db	808	CCGCGACA	CCGAGGGACTTCTTATGCGGACATCATCTTCGGCAATGATCTCGATGCC		867
Qy	862	AAGCGCGCG	CGCATTCGCGCGGCGCATATTTCAAGCGCGCGGACGTAACGCGGCTGTGCTG		921
Db	868	AAGCGCGAG	CCGATCCCGCGGACATTTATGCACTGCCATGTATCACTCGCTGCTGTT		927
Qy	922	GATGACGT	CTCGGCCCAAGCGCTGTGACGCTTGATGCCGATTCGAACTCGCAAAACGAG		981
Db	928	AACAATG	CGCCGGAACCGCGTGAAGAGCTGGCGTTCGCGTGCATGACGAGGCGGAA		987
Qy	982	G	982		
Db	988	G	988		

RESULT 13  
AP006618\_34  
WPCOMMENT

Fragment Name	Begin	End	LOCUS	AP006618	Accession	AP006618
AP006618_00	1	110000				
AP006618_01	100001	210000				
AP006618_02	200001	310000				
AP006618_03	300001	410000				
AP006618_04	400001	510000				
AP006618_05	500001	610000				
AP006618_06	600001	710000				



AP006618_07	700001	810000
AP006618_08	800001	910000
AP006618_09	900001	1010000
AP006618_10	1000001	1110000
AP006618_11	1100001	1210000
AP006618_12	1200001	1310000
AP006618_13	1300001	1410000
AP006618_14	1400001	1510000
AP006618_15	1500001	1610000
AP006618_16	1600001	1710000
AP006618_17	1700001	1810000
AP006618_18	1800001	1910000
AP006618_19	1900001	2010000
AP006618_20	2000001	2110000
AP006618_21	2100001	2210000
AP006618_22	2200001	2310000
AP006618_23	2300001	2410000
AP006618_24	2400001	2510000
AP006618_25	2500001	2610000
AP006618_26	2600001	2710000
AP006618_27	2700001	2810000
AP006618_28	2800001	2910000
AP006618_29	2900001	3010000
AP006618_30	3000001	3110000
AP006618_31	3100001	3210000
AP006618_32	3200001	3310000
AP006618_33	3300001	3410000
AP006618_34	3400001	3510000
AP006618_35	3500001	3610000
AP006618_36	3600001	3710000
AP006618_37	3700001	3810000
AP006618_38	3800001	3910000
AP006618_39	3900001	4010000
AP006618_40	4000001	4110000
AP006618_41	4100001	4210000
AP006618_42	4200001	4310000
AP006618_43	4300001	4410000
AP006618_44	4400001	4510000
AP006618_45	4500001	4610000
AP006618_46	4600001	4710000
AP006618_47	4700001	4810000
AP006618_48	4800001	4910000
AP006618_49	4900001	5010000
AP006618_50	5000001	5110000
AP006618_51	5100001	5210000
AP006618_52	5200001	5310000
AP006618_53	5300001	5410000
AP006618_54	5400001	5510000
AP006618_55	5500001	5610000
AP006618_56	5600001	5710000
AP006618_57	5700001	5810000
AP006618_58	5800001	5910000
AP006618_59	5900001	6010000
AP006618_60	6000001	601225

	Query Match	Similarity	42.7%	Score 445;	DB 1:	Length 110000;
	Best Local	Similarity	68.1%;	Pred. No. 2.4e-48;		
	Matches	619;	Conservative	0;	Mismatches 290;	Indels 0; Gaps 0;
QY	25	CGCGGCGGGCGGTGCAGGCCGCGCGGGTGTTCCTCGATCTCGACCCGACAGTCGAGAA	84			
Db	70070	CGCCCGCGGGCGGTGCAGGCCGCGCGGGTGTGCTGCACGCGCGCGCAGCGGTGCACAG	70129			
QY	85	GCGATTGCGCCCTGATTCGAGCAGCGCGCCGACAGAGAAGTGCGCCCTGATTCGCAATTTCCCAAG	144			
Db	70130	TGCGTTCGGGTTCGATTCGAGAGAGCGCACCAACGAGGCGGCGTGAATTCGCTTCCCGAAG	70189			
QY	145	ACTTGATATCCCGGCTATCCCTTTGATATAGCTGAGCGCGCCGAGCTGGGGCATCGC	204			
Db	70190	ACGTTTCGAGCCCGGGTACC CGTGATGGCTGTGCTGATTTGCGCGGCTGGGGCATTCAG	70249			
QY	205	TTTCGTCACGCGCTATTTTCGAGAAATTCGCTCGTCGCGCGGACGAAAGATGGCAGCGCCCTG	264			

Db	70250	TTTCGTGGCCCGGCTACTTTCGAACTCGCTTCGCGTGAACGGCCCGCTTGTTCGCCCGGTTG	70305
Qy	265	GGGATGCGGCGCCCGCGCCGACCGGCATGATGTCGTGACCCTGATTAAGCAGCGCGCGGCG	324
Db	70310	CGCAGAGCGCGCCCGCCCGACCGGATCAACGTGTGTGATACCGGGGACAGCGACGGCGACGCGC	70365
Qy	325	GGCAGCCTCTATATATGAGGCGCAGGCGATCTTTGGGCGCCCGCATATGGCATCTGATTCGCGCGCGC	384
Db	70370	GGTTCCGCTGTACATATGGGCGCAGGGGATCATATCGGGGGCCGACGGCGAGGTGCTCCCGCGCCCT	70429
Qy	385	CGCAAGCTCAAGCGCTTACCCATGCGGAGCGGACCGTGTTCGGCGAGGAGAGACGGCAACCAT	444
Db	70430	CGAAGGCTCAAGCGCGACCGACCGTCGAGCGGACCGCTTTCGGTGTAGAGACACGCTCGAAC	70489
Qy	445	CTCCGCGTGTACATACGATACCGGCATCGGAGCGCTCGGCGCGCTCTGTTGTCTGGAGACATTC	504
Db	70490	CTGACGCGTGTGTGTGACACCGAGCTGGGCGCGCTTCGGCGCCCTGTGTCTGTCTGGAGCATCTG	70549
Qy	505	CAGCCATTGTTCGAAATACCGCCATGTACGCCGCGGACGAAACAGGTCCAGTCCGCTCGTGG	564
Db	70550	CAGCGCGTGTACCAAGTACCGGATGTACTCCAGACAGACAAATCCACCTGTGCGCGCTGG	70609
Qy	565	CCGAGCTTTCAGCCTCTATCGCGGCATAGGCTTATGCGCTCGAGACCGGAGGTCAATACCGCC	624
Db	70610	CCGAGCTTTCGCGTCTACCGGGCGCGCGGTACGCCCTTGCGCCCGAGGTCAACACCGCG	70665
Qy	625	GCAAGCCAGATCTAACCGCGGTTCAGAGGCGCGCTGTAACTGTCTGAGCGTCTGTGCGGCACCTT	684
Db	70670	GCGGCGCGGACGATACGGGTCGAGGGCGACGTGTCTGTGTCTCTCGCGCGGCGCGGTATTC	70729
Qy	685	TCGCGCGGAGATGATTCAGAGTATTGTGTGTGTATACGCCCGACAGAGATGTTCTTCAAGGCC	744
Db	70730	GACGAGGCGCGGTGTGCAATGTGTTCTCGACACACCGGCGCAAGCGGCAATCTGCTGCTGCC	70789
Qy	745	GGCGCGCGTTCCTTCACATGATTTTCGGGCGCCGACAGCGCGCGCCCTGGCCGACGCGCTCCCG	804
Db	70790	GGCGCGCGCTTCGCGCCAGATCTACGGCGCGGAGCGGAGGCTGGGACACGGCGTTGCCG	70849
Qy	805	GAGACCGAGAGGAGCTGCTGTGTGCGCGCATATCGACCTTGCGGATGATTCGCTGTGGCCAG	864
Db	70850	GAGACCGAGAGGAGGCGCTGTGTCTACGCGCATTCGAGGCGCTCGGCGCGGTGCGCAAG	70909
Qy	865	GCGCGCGCGCGATCCGCGCGGCGCCATTCATTCACGCGCCCGGACGTTAACGGCGTGTCTCAT	924
Db	70910	AGCGCGGCGCACCGCGGTGCGGACCTACAGCGCGCCGAGCGTGTCTGCACTGTGTGGAT	70969
Qy	925	CGACGTCGCG 933	
Db	70970	CCGCGGCGCG 70978	
RESULT 14			
LOCUS	AY487448	1017 bp	DNA linear ENV 05-APR-2004
DEFINITION	Uncultured organism clone 2A25 nitrilase (BD513) gene, complete cds.		
ACCESSION	AY487448		
VERSION	AY487448.1	GI:40890100	
KEYWORDS	ENV.		
SOURCE	uncultured organism		
ORGANISM	unclassified; environmental samples.		
REFERENCE	1 (bases 1 to 1017)		
AUTHORS	Robertson,D.E., Chaplin,J.A., Desantis,G., Podar,M., Madden,M., Chi,E., Richardson,T., Milan,A., Miller,M., Weiner,D.P., Wong,K., McQuaid,J., Farwell,B., Preston,L.A., Tan,X., Sneed,M.A., Keller,M., Mathur,E., Kretz,P.L., Burk,M.J. and Short,J.M.		
TITLE	Exploring nitrilase sequence space for enantioselective catalysis		
JOURNAL	Appl. Environ. Microbiol. 70 (4), 2429-2436 (2004)		
REFERENCE	2 (bases 1 to 1017)		
AUTHORS	Robertson,D.E., Chaplin,J.A., Desantis,G., Podar,M., Madden,M., Chi,E., Richardson,T.H., Milan,A., Miller,M., Weiner,D.P., Wong,K.,		







GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

## OM nucleic - nucleic search, using sw model

Run on: April 26, 2006, 22:18:14 ; Search time 719.835 Seconds

(without alignments)  
9638.243 Million cell updates/sec

Title: US-09-751-299-1

Perfect score: 1041  
Sequence: 1 atgcgcgagcccatgaacgaa.....gcgcgcgcgcgcgcgcagtag 1041Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Database : N\_Geneseq 21:\*

1: geneseqn1980s:\*  
2: geneseqn1990s:\*  
3: geneseqn2000s:\*  
4: geneseqn2001as:\*  
5: geneseqn2001bs:\*  
6: geneseqn2002as:\*  
7: geneseqn2002bs:\*  
8: geneseqn2003as:\*  
9: geneseqn2003bs:\*  
10: geneseqn2003ds:\*  
11: geneseqn2003as:\*  
12: geneseqn2004as:\*  
13: geneseqn2004bs:\*  
14: geneseqn2005s:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1041	100.0	1041	4 AAD11166	Aad11166 Nitriase
2	1041	100.0	1041	10 ADC24116	Adc24116 DNA seque
3	1041	100.0	1041	10 ADH36217	Adh36217 Chemical
4	1041	100.0	1041	12 ADG93916	Adg93916 Nitriase
5	1041	100.0	1041	12 ADI62515	Adi62515 DNA encod
6	1041	100.0	1041	12 ADI64636	Adi64636 DNA encod
7	682	65.5	1026	10 ADC23766	Adc23766 DNA seque
8	682	65.5	1026	12 ADH35867	Adh35867 Chemical
9	682	65.5	1026	12 ADG93568	Adg93568 Nitriase
10	682	65.5	1026	12 ADI62165	Adi62165 DNA encod
11	682	65.5	1026	12 ADI64286	Adi64286 DNA encod
12	486	46.7	1062	10 ADC23944	Adc23944 DNA seque
13	486	46.7	1062	12 ADH36045	Adh36045 Chemical
14	486	46.7	1062	12 ADG93746	Adg93746 Nitriase
15	486	46.7	1062	12 ADI62343	Adi62343 DNA encod
16	486	46.7	1062	12 ADI64464	Adi64464 DNA encod
17	478.2	45.9	1056	10 ADC24074	Adc24074 DNA seque
18	478.2	45.9	1056	12 ADH36175	Adh36175 Chemical
19	478.2	45.9	1056	12 ADG93875	Adg93875 Nitriase

20	478.2	45.9	1056	12 ADI62473	Adi62473 DNA encod
21	478.2	45.9	1056	12 ADI64594	Adi64594 DNA encod
22	470	45.1	1050	10 ADC24110	Adc24110 DNA seque
23	470	45.1	1050	12 ADH36211	Adh36211 Chemical
24	470	45.1	1050	12 ADG93910	Adg93910 Nitriase
25	470	45.1	1050	12 ADI62509	Adi62509 DNA encod
26	470	45.1	1050	12 ADI64630	Adi64630 DNA encod
27	457.2	43.9	1065	10 ADC23834	Adc23834 DNA seque
28	457.2	43.9	1065	12 ADH35935	Adh35935 Chemical
29	457.2	43.9	1065	12 ADG93636	Adg93636 Nitriase
30	457.2	43.9	1065	12 ADI62233	Adi62233 DNA encod
31	457.2	43.9	1065	12 ADI64354	Adi64354 DNA encod
32	455.2	43.7	1077	10 ADC23902	Adc23902 DNA seque
33	455.2	43.7	1077	12 ADH36003	Adh36003 Chemical
34	455.2	43.7	1077	12 ADG93704	Adg93704 Nitriase
35	455.2	43.7	1077	12 ADI62301	Adi62301 DNA encod
36	455.2	43.7	1077	12 ADI64422	Adi64422 DNA encod
37	454.6	43.7	1077	10 ADC23930	Adc23930 DNA seque
38	454.6	43.7	1077	12 ADH36031	Adh36031 Chemical
39	454.6	43.7	1077	12 ADG93732	Adg93732 Nitriase
40	454.6	43.7	1077	12 ADI62329	Adi62329 DNA encod
41	454.6	43.7	1077	12 ADI64450	Adi64450 DNA encod
42	453.6	43.6	1017	10 ADC23982	Adc23982 DNA seque
43	453.6	43.6	1017	12 ADH36083	Adh36083 Chemical
44	453.6	43.6	1017	12 ADG93784	Adg93784 Nitriase
45	453.6	43.6	1017	12 ADI62381	Adi62381 DNA encod

## ALIGNMENTS

RESULT 1	
AD11166	AD11166 standard; DNA, 1041 BP.
ID	AD11166
AC	AD11166;
XX	
DT	25-SEP-2001 (first entry)
XX	
DE	Nitriase DNA #1.
XX	
KW	Nitriase; enantiomer; alpha-substituted carboxylic acid;
KW	stereoselective; hydrolysis; amino nitriase; cyanohydrin; de.
OS	Unidentified.
XX	
FT	Key
FT	Location/Qualifiers
FT	1..1041
FT	/*tag= a
FT	/product= "Nitriase protein #1"
XX	
PD	05-JUL-2001.
XX	
XX	W0200148175-A2.
XX	
XX	29-DEC-2000; 2000MO-US035555.
XX	
XX	29-DEC-1999; 99US-0173609P.
XX	
XX	07-DEC-2000; 2000US-0254414P.
XX	
XX	(DIVE-) DIVERSA CORP.
XX	(MADN/) MADNEN D.
XX	
XX	Madden M, Weiner DP, Chaplin JA;
XX	
XX	WPI: 2001-465211/50.
XX	
XX	P-PSDB; AA05488.
XX	
XX	Producing alpha-substituted carboxylic acid enantiomers by contacting
XX	aldehyde or ketone with cyanide, ammonia compound or its salt or amine
XX	PT and hydrolyzing the resulting amino nitriase or cyanohydrin with
XX	Nitriase.

PS Claim 27; Page 82-84; 87pp; English.

XX The present invention relates to methods for producing an  
CC enantiomerically pure alpha-substituted carboxylic acid. The method  
CC involves contacting an aldehyde or ketone with a cyanide containing  
CC compound and an ammonia-containing compound or an ammonium salt or amine,  
CC and stereoselectively hydrolysing the resulting amino nitrile or  
CC cyanohydrin intermediate with a nitrilase or a polypeptide having  
CC nitrilase activity. The present sequence is a DNA encoding nitrilase #1  
XX  
SQ Sequence 1041 BP; 172 A; 340 C; 357 G; 172 T; 0 U; 0 Other;

Query Match 100.0%; Score 1041; DB 4; Length 1041;  
Best Local Similarity 100.0%; Pred. No. 1,1e-197;  
Matches 1041; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 ATGTGGAGCCCATGACGAAGTATCGCGCGCGCGGTGACAGCGCGCGGTGCTTC 60  
DB 1 ATGTGGAGCCCATGACGAAGTATCGCGCGCGCGGTGACAGCGCGCGGTGCTTC 60  
QY 61 GATCTCGACCGCAGTCAGTGAAGAGCATCGGCTGATTCAGCAGCGCGCCAGCAGAC 120  
DB 61 GATCTCGACCGCAGTCAGTGAAGAGCATCGGCTGATTCAGCAGCGCGCCAGCAGAC 120  
QY 121 GTGCGCTGATCGCATTCGAGACTTGGATTCCCGCTATCCCTTTGGATATGCTG 180  
DB 121 GTGCGCTGATCGCATTCGAGACTTGGATTCCCGCTATCCCTTTGGATATGCTG 180  
QY 181 GGGCGCGCGGCTTGGGGCATGCGCTTGTCAGAGCGTATTCAGAAATTCGCTCGTCCG 240  
DB 181 GGGCGCGCGGCTTGGGGCATGCGCTTGTCAGAGCGTATTCAGAAATTCGCTCGTCCG 240  
QY 241 GGCAGCAGAGAGTGGCAGCGCTTGCGGATGCGCGCGCGCCAGCAGCATATGTCGTG 300  
DB 241 GGCAGCAGAGAGTGGCAGCGCTTGCGGATGCGCGCGCGCCAGCAGCATATGTCGTG 300  
QY 301 GCGCGCTATAGCCAGCGCGCGCGCGCGAGCCTCTATATGGGCCAGCGCATCTTCGCGCCC 360  
DB 301 GCGCGCTATAGCCAGCGCGCGCGCGCGAGCCTCTATATGGGCCAGCGCATCTTCGCGCCC 360  
QY 361 GATGCGCATGATCGCGCGCGCGCGCGCAAGCTCAAGCTCAAGTCGCGAGCGCAGCGT 420  
DB 361 GATGCGCATGATCGCGCGCGCGCGCGCAAGCTCAAGCTCAAGTCGCGAGCGCAGCGT 420  
QY 421 TTGCGCGAGGAGACGCGCAGCCATCTTCGCGTGCAGATACCGCATCGCGCGCTCGGC 480  
DB 421 TTGCGCGAGGAGACGCGCAGCCATCTTCGCGTGCAGATACCGCATCGCGCGCTCGGC 480  
QY 481 GCGCTCTGTGCTGGGAGCATTCGAGCATTTGTCGAATAACGCCATGTAACGCCCGAG 540  
DB 481 GCGCTCTGTGCTGGGAGCATTCGAGCATTTGTCGAATAACGCCATGTAACGCCCGAG 540  
QY 541 GAAACAGTCCAGCGTGGCGTGTGGCCGAGCTTCAAGCTCTATCGCGCATGCGCTATGCG 600  
DB 541 GAAACAGTCCAGCGTGGCGTGTGGCCGAGCTTCAAGCTCTATCGCGCATGCGCTATGCG 600  
QY 601 CTGCGACCGGAGGTCAATACCGCGCGAGCATCTACCGCGGTTCGAGGCGCGCTGTAC 660  
DB 601 CTGCGACCGGAGGTCAATACCGCGCGAGCATCTACCGCGGTTCGAGGCGCGCTGTAC 660  
QY 661 GTGCTGGCGCTGTCGCGGACCGTTCGCGCGAGATGATCAAGGTATTTGGATACGCC 720  
DB 661 GTGCTGGCGCTGTCGCGGACCGTTCGCGCGAGATGATCAAGGTATTTGGATACGCC 720  
QY 721 GACAAGAGATGTTCTCAAGCGCGCGCGGCTTTGGCATGATTTTGGGCGCCGAGCG 780  
DB 721 GACAAGAGATGTTCTCAAGCGCGCGCGGCTTTGGCATGATTTTGGGCGCCGAGCG 780  
QY 781 CGCGCCCTGGCGAGCGCGCTCCGAGAGACCGAAGAGGAGTCTGCTGCGCGATATCGAC 840  
DB 781 CGCGCCCTGGCGAGCGCGCTCCGAGAGACCGAAGAGGAGTCTGCTGCGCGATATCGAC 840  
QY 841 CTCGGCATGATCGCGCTTGGCCAAAGCGCGCGCATTCGCGCGGCGCATATTCAGGCGCC 900

DB 841 CTCGGCATGATCGCGCTTGGCCAAAGCGCGCGCATTCGCGCGGCGCATATTCAGGCGCC 900  
QY 901 GACGTAAACGGGCTGCTGCTGATTCAGTCCGAGCCCAACGCGTCTGACGCTGATGCC 960  
DB 901 GACGTAAACGGGCTGCTGCTGATTCAGTCCGAGCCCAACGCGTCTGACGCTGATGCC 960  
QY 961 GCATTGCAACCGCAAAACGAGGACCAAGGCGGACGCGCGCTGCGCGTGTGGCGGAA 1020  
DB 961 GCATTGCAACCGCAAAACGAGGACCAAGGCGGACGCGCGCTGCGCGTGTGGCGGAA 1020  
QY 1021 AGCGCGCGCGCGCGCATGTAG 1041  
DB 1021 AGCGCGCGCGCGCGCATGTAG 1041

RESULT 2  
ADCC4116  
ID ADC24116 standard; DNA; 1041 BP.  
XX  
AC ADC24116;  
XX  
DT 18-DEC-2003 (first entry)  
XX  
DE DNA sequence (SeqID 383) encoding a nitrilase enzyme.  
XX  
KW gene; ds; nitrilase; nitrile; cyanohydrin; ammonia; biocatalyst;  
XX  
OS unidentified.  
XX  
PN WO2003000840-A2.  
XX  
PD 03-JAN-2003.  
XX  
PF 15-MAY-2002; 2002WO-US015983.  
XX  
PR 21-JUN-2001; 2001US-0300189P.  
XX  
PR 30-JUL-2001; 2001US-030906P.  
XX  
PR 22-JAN-2002; 2002US-0351336P.  
XX  
PA (DIVE-) DIVERSA CORP.  
PA (MADD) MADDEN D.  
XX  
PI Madden M, Desantis G, Chaplin JA, Weiner DP, Milan A, Chi E;  
PI Short JM, Burk M;  
XX  
DR WPI: 2003-201417/19.  
XX  
PT P-PSDB; ADC24117.  
XX  
PT Novel nitrilase polypeptide, useful for making (R)- or (S)-ethyl-4-cyano-  
PT 3-hydroxybutyric acid or (R)- or (S)-mandelic acid or (S)- or (R)-phenyl  
PT lactic acid derivative and for producing pharmaceutical composition, and  
PT food additive.  
XX  
PS Claim 1; SEQ ID NO 383; 560pp; English.  
XX  
XX This invention relates to nitrilases and the nucleic acids that encode  
XX these enzymes thereof. Specifically, it refers to polypeptides that  
XX exhibit nitrilase activity, i.e. the ability to directly hydrolyse  
XX nitriles or cyanohydrins into their corresponding carboxylic acids and  
XX ammonia. Nitrilases have commercial utility as biocatalysts for use in  
XX the synthesis of enantiomerically pure aromatic and aliphatic amino  
XX acids, as well as hydroxy acids, which are important for the development  
XX of chiral medicines. Furthermore, the present invention describes  
XX nitrilases, isolated from mesophilic microorganisms, that have improved  
XX activity and stability at increased pH and temperature. They are also  
XX inexpensive, efficient catalysts, have broad substrate specificity and  
XX are capable of chiral differentiation. This polypeptide is a DNA  
XX sequence that encodes a nitrilase enzyme of the invention.  
SQ Sequence 1041 BP; 172 A; 340 C; 357 G; 172 T; 0 U; 0 Other;

Query Match 100.0%; Score 1041; DB 10; Length 1041;  
Best Local Similarity 100.0%; Pred. No. 1.1e-197;  
Matches 1041; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 ATGTGGAGCCCATGACGAAGTATCGCGCGCGCGGCTGACGCGCGCGGTTCTTC 60
DB 1 ATGTGGAGCCCATGACGAAGTATCGCGCGCGCGGCTGACGCGCGCGGTTCTTC 60
QY 61 GATCTGACCGCGACAGTCGAGAAAGCATCGGCTGTATCCAGAGCGCGCGCAAGAGAC 120
DB 61 GATCTGACCGCGACAGTCGAGAAAGCATCGGCTGTATCCAGAGCGCGCGCAAGAGAC 120
QY 121 GTGCGCTGATCGCATTCGACAGACTTGATTCGCGCTATTCCTTTGGATATGCTG 180
DB 121 GTGCGCTGATCGCATTCGACAGACTTGATTCGCGCTATTCCTTTGGATATGCTG 180
QY 181 GCGCGCGCGGCTTGGGCGATCGCTTCGTCCAGGCTATTTGAGAAATTCGCTGTCGC 240
DB 181 GCGCGCGCGGCTTGGGCGATCGCTTCGTCCAGGCTATTTGAGAAATTCGCTGTCGC 240
QY 241 GGCAGCAGACAGATGCGAGCGCTTGGCGATGCGGCGCGCGCGCGCATATGCTGTCG 300
DB 241 GGCAGCAGACAGATGCGAGCGCTTGGCGATGCGGCGCGCGCGCGCATATGCTGTCG 300
QY 301 GCGCGCTATAGCGAGCGCGCGCGCGCGCGCTATATAGGCGCGAGATTTTCGCGCC 360
DB 301 GCGCGCTATAGCGAGCGCGCGCGCGCGCGCTATATAGGCGCGAGATTTTCGCGCC 360
QY 361 GATGCGCATGTGATTCGCGCGCGCGCGCGAGCTCAAGCTCAAGCTCAAGCTCAAGCT 420
DB 361 GATGCGCATGTGATTCGCGCGCGCGCGCGAGCTCAAGCTCAAGCTCAAGCTCAAGCT 420
QY 421 TTGCGCGAGGAGAGCGCGCGCATCTTCGCGTGCAGATACCGCATCGCGCGCTTCGCG 480
DB 421 TTGCGCGAGGAGAGCGCGCGCATCTTCGCGTGCAGATACCGCATCGCGCGCTTCGCG 480
QY 481 GCGCTCTGTTGCTGGAGGACATCGACCATATGTCGAAATACGCGCATATGCGCGCGAC 540
DB 481 GCGCTCTGTTGCTGGAGGACATCGACCATATGTCGAAATACGCGCATATGCGCGCGAC 540
QY 541 GAAAGGTTCACGTCGCGCGTGCAGCTTCAGCTCTATCGCGCATATGCGCTATGCG 600
DB 541 GAAAGGTTCACGTCGCGCGTGCAGCTTCAGCTCTATCGCGCATATGCGCTATGCG 600
QY 601 CTGCGACCGGAGGTCAATACCGCGCGCAAGCCAGATCTACGCGGTCGAGCGGCTGTAC 660
DB 601 CTGCGACCGGAGGTCAATACCGCGCGCAAGCCAGATCTACGCGGTCGAGCGGCTGTAC 660
QY 661 GTGCTGCGTGTGCGCGGACCGTTTCGCGGAGATGATCAAGGATATGTCGATACGCC 720
DB 661 GTGCTGCGTGTGCGCGGACCGTTTCGCGGAGATGATCAAGGATATGTCGATACGCC 720
QY 721 GACAAGAGATGTTCTCAAGGCGCGCGGTTTTCATGATTTTCGAGCGCGAGCGG 780
DB 721 GACAAGAGATGTTCTCAAGGCGCGCGGTTTTCATGATTTTCGAGCGCGAGCGG 780
QY 781 CGCGCGCTGCGCGAGCGCTTCGCGAGACCGAAGAGGAGCTGCTGAGTCGCGCATATGAC 840
DB 781 CGCGCGCTGCGCGAGCGCTTCGCGAGACCGAAGAGGAGCTGCTGAGTCGCGCATATGAC 840
QY 841 CTGCGCGCATGATCGCTTGGCGCAAGGCGCGCGCGATCCGCGCGCGCATATTCACGCGCC 900
DB 841 CTGCGCGCATGATCGCTTGGCGCAAGGCGCGCGCGATCCGCGCGCGCATATTCACGCGCC 900
QY 901 GACGTAAACGCGGCTGCTGCTGATCGATCGCGCGCAAGCGCTGTCAGCTTTGATGCG 960
DB 901 GACGTAAACGCGGCTGCTGCTGATCGATCGCGCGCAAGCGCTGTCAGCTTTGATGCG 960
QY 961 GCATTGCAACCGCAAAACGAGACAGAGCGCGCGCGCTGCGCGCTGTCGTCGCGAA 1020
DB 961 GCATTGCAACCGCAAAACGAGACAGAGCGCGCGCGCTGCGCGCTGTCGTCGCGAA 1020
QY 1021 AGCGCGCGCGCGCGCGCATAG 1041
```

DB 1021 AGCGCGCGCGCGCGCATAG 1041

```
RESULT 3
ADH36217
ID ADH36217 standard; DNA; 1041 BP.
XX
AC ADH36217;
XX
DT 11-MAR-2004 (first entry)
XX
DE Chemical process monitoring-related nitrilase gene sequence SeqID383.
XX
KM chemical process monitoring; biochemical process monitoring; cyanide;
XX high throughput system; gene; db.
XX
OS Unidentified.
XX
PN W02003098187-A2.
XX
PD 27-NOV-2003.
XX
PF 15-MAY-2003; 2003MO-US015639.
XX
PR 15-MAY-2002; 2002US-0380737P.
XX
PA (DIVE-) DIVERSA CORP.
XX
PI Weiner D, Chaplin JA, Chi E, Milan A, Desantis G, Burk MJ;
PI Meguaid J, Stege J;
XX
DR WPI; 2004-142708/14.
XX
DR P-PSDB; ADH36218.
XX
PT Monitoring a chemical or biochemical process comprises providing a
PT reactant comprising a cyanide or a material that can be converted to
PT cyanide or a reactant that generates a cyanide or a material that can be
PT converted to cyanide.
XX
XX Claim 74; SEQ ID NO 383; 277bp; English.
XX
PS This invention relates to a novel method of monitoring chemical or
PS biochemical processes. The method involves providing a reactant
CC comprising cyanide (or a material that can be converted to a cyanide)
CC that generates as a reaction product cyanide or a material that can be
CC converted to cyanide and measuring the concentration of produced cyanide.
CC The method is useful for monitoring a chemical or biochemical process.
CC The method is effective for high throughput systems and is sufficiently
CC sensitive to detect a small amount of product. The present sequence is
CC that of a gene which encodes a nitrilase enzyme which can be used in the
CC method of the invention.
XX
SQ Sequence 1041 BP; 172 A; 340 C; 357 G; 172 T; 0 U; 0 Other:
Query Match 100.0%; Score 1041; DB 12; Length 1041;
Best Local Similarity 100.0%; Pred. No. 1.1e-197;
Matches 1041; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 ATGTGGAGCCCATGACGAAGTATCGCGCGCGCGGCTGACGCGCGCGGTTCTTC 60
DB 1 ATGTGGAGCCCATGACGAAGTATCGCGCGCGCGGCTGACGCGCGCGGTTCTTC 60
QY 61 GATCTGACCGCGACAGTCGAGAAAGCATCGGCTGTATCCAGAGCGCGCGCAAGAGAC 120
DB 61 GATCTGACCGCGACAGTCGAGAAAGCATCGGCTGTATCCAGAGCGCGCGCAAGAGAC 120
QY 121 GTGCGCTGATCGCATTCGACAGACTTGATTCGCGCTATTCCTTTGGATATGCTG 180
DB 121 GTGCGCTGATCGCATTCGACAGACTTGATTCGCGCTATTCCTTTGGATATGCTG 180
QY 181 GCGCGCGCGGCTTGGGCGATCGCTTCGTCCAGGCTATTTGAGAAATTCGCTGTCGC 240
DB 181 GCGCGCGCGGCTTGGGCGATCGCTTCGTCCAGGCTATTTGAGAAATTCGCTGTCGC 240
```

```
Db 181 GCGCGCGCGGCTTGGGGCATGCGCTTCGTCACAGCGCTATTTCGAAATTGCTCGTCGCGC 240
QY 241 GCGCAGCAGCAGTGGCAGAGCCCTTGGCGGATGCGGCCGCCGCA CGGCATGATGCTGTG 300
Db 241 GCGCAGCAGCAGTGGCAGAGCCCTTGGCGGATGCGGCCGCCGCA CGGCATGATGCTGTG 300
QY 301 GCCGCTATAGCGAGCGCGCGGGGAGAGCCTTATATAGGCGCGAGGATTTGGGCC 360
Db 301 GCCGCTATAGCGAGCGCGCGGGGAGAGCCTTATATAGGCGCGAGGATTTGGGCC 360
QY 361 GATGCGCATCTGATTCGCGCGCGCGCAAGCTTAAGCTTACCATGCGAGCGACCGTG 420
Db 361 GATGCGCATCTGATTCGCGCGCGCGCAAGCTTACCATGCGAGCGACCGTG 420
QY 421 TTGCGCGAGGAGAGCGAGCGCATCTGCGCGTGAAGATACCGGCATTCGGCGCTCGGC 480
Db 421 TTGCGCGAGGAGAGCGAGCGCATCTGCGCGTGAAGATACCGGCATTCGGCGCTCGGC 480
QY 481 GCGCTCTGTGCTGGAGCAGCATCAGCCATTTGTCGAATATCGCCATGTACGCCCGAC 540
Db 481 GCGCTCTGTGCTGGAGCAGCATCAGCCATTTGTCGAATATCGCCATGTACGCCCGAC 540
QY 541 GAATAGGTCCACGTCGCGTGTGCGCGAGCTTCAGCTTCTATCGCGCATGCGCTATGCG 600
Db 541 GAATAGGTCCACGTCGCGTGTGCGCGAGCTTCAGCTTCTATCGCGCATGCGCTATGCG 600
QY 601 CTGCGACCGAGATCAATACCGCGCGCAAGCAGATCTTACGCGGTCAGGCGCGCTGTAC 660
Db 601 CTGCGACCGAGATCAATACCGCGCGCAAGCAGATCTTACGCGGTCAGGCGCGCTGTAC 660
QY 661 GTGCTGCGGTGTCGCGCGAGCGTTTGGCGCGAGATGATCAAGGATTTGATGATCGCC 720
Db 661 GTGCTGCGGTGTCGCGCGAGCGTTTGGCGCGAGATGATCAAGGATTTGATGATCGCC 720
QY 721 GACAAGAGATGTTCTTCAAGCGCGCGCGGCTTTTGCATGATTTTCGCGCGCGAGCGC 780
Db 721 GACAAGAGATGTTCTTCAAGCGCGCGCGGCTTTTGCATGATTTTCGCGCGCGAGCGC 780
QY 781 CCGCGCTGCGCGAGCGCGCTTCCGCGAGACCGAAGAGGAGATGTCGTCGCGATATCGAC 840
Db 781 CCGCGCTGCGCGAGCGCGCTTCCGCGAGACCGAAGAGGAGATGTCGTCGCGATATCGAC 840
QY 841 CTGCGCATGATCGGTCGCGCGCAAGCGCGCGCATCGCGCGCGCACTATTACGCGCC 900
Db 841 CTGCGCATGATCGGTCGCGCGCAAGCGCGCGCATCGCGCGCGCACTATTACGCGCC 900
QY 901 GACGTAAACGCGGCTGCTGCTGATTCGACGTCGCGCCCAACGCGTCTGACGCTTGATGCC 960
Db 901 GACGTAAACGCGGCTGCTGCTGATTCGACGTCGCGCCCAACGCGTCTGACGCTTGATGCC 960
QY 961 GCATTGCGAACCAGAAACAGAGACAGAGGCGAGCGCGCGCGCTGCGCGTGTGGCGAA 1020
Db 961 GCATTGCGAACCAGAAACAGAGACAGAGGCGAGCGCGCGCGCTGCGCGTGTGGCGAA 1020
QY 1021 AGCGCGCGCGCGCGCGCAGTAG 1041
Db 1021 AGCGCGCGCGCGCGCAGTAG 1041

RESULT 4
ADG93916
ID ADG93916 standard; DNA; 1041 BP.
XX
XX ADG93916;
XX
XX 11-MAR-2004 (first entry)
XX
XX Nitrilase enzyme gene sequence SegID383.
XX
XX nitrilase; nitrile; carboxylic acid; chemical process; pH; temperature;
XX enantioselective transformation; gene; ds.
XX
XX Unidentified.
```

```
XX
XX WO2003097810-A2.
XX
XX 27-NOV-2003.
XX
XX 15-MAY-2003; 2003WO-US015712.
XX
XX 15-MAY-2002; 2002US-00146772.
XX
XX 09-SEP-2002; 2002US-00241742.
XX
XX (DIVE-) DIVERSA CORP.
XX
XX Desantis G, Short JM, Burk M, Wong K, Farwell R, Chatman K;
XX WPI; 2004-090637/09.
XX
XX P-PDB; ADG93917.
XX
XX New isolated or recombinant nucleic acid encoding a polypeptide having
XX nitrilase activity, useful for screening enantioselective transformation.
XX
XX Claim 44; SEQ ID NO 383; 295bp; English.
XX
XX This invention is related to a novel isolated or recombinant nucleic acid
XX encoding a protein having nitrilase activity. Nitrilase's are capable of
XX converting nitrile's directly to carboxylic acids and have great
XX potential for use in industrial chemical processes. The isolated
XX nitrilase proteins of the invention have increased activity and stability
XX at increased pH and temperature when compared to those conventionally
XX used. In addition, the nucleic acid of the invention is useful for
XX screening enantioselective transformation. The present sequence is that
XX of a DNA sequence which encodes a nitrilase enzyme of the invention.
XX
XX Sequence 1041 BP; 172 A; 340 C; 357 G; 172 T; 0 U; 0 Other;
XX
XX Query Match 100.0%; Score 1041; DB 12; Length 1041;
XX Best Local Similarity 100.0%; Pred. No. 1.1e-197;
XX Matches 1041; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX 1 ATGTCGAGCCCATGACGAAATATCGCGCGCGCGGTGACAGCGCGCGGCTGTTCTC 60
XX 1 ATGTCGAGCCCATGACGAAATATCGCGCGCGCGGTGACAGCGCGCGGCTGTTCTC 60
XX
XX 61 GATCTGACCGGACAGTGGAGAAAGGATCGGCGCTGATGAGACAGCGCGCGCAAGCAGAGC 120
XX 61 GATCTGACCGGACAGTGGAGAAAGGATCGGCGCTGATGAGACAGCGCGCGCAAGCAGAGC 120
XX
XX 121 GTGCGCTGATGCAATTCGACAGACTTGATTTCCGCGCTATCCCTTTGATATGGCTG 180
XX 121 GTGCGCTGATGCAATTCGACAGACTTGATTTCCGCGCTATCCCTTTGATATGGCTG 180
XX
XX 181 GCGCGCGCGGCTTGGGGCATGCGCTTCGTCAGCGCTATTTGAGAAATTCGTCGCGC 240
XX 181 GCGCGCGCGGCTTGGGGCATGCGCTTCGTCAGCGCTATTTGAGAAATTCGTCGCGC 240
XX
XX 241 GCGCAGCAGCAGTGGAGGCGCTTGGCGGATGCGCGCGCGCGCAGCGCATGATGCTGTG 300
XX 241 GCGCAGCAGCAGTGGAGGCGCTTGGCGGATGCGCGCGCGCGCAGCGCATGATGCTGTG 300
XX
XX 301 GCGCGCTATAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCATCTTTCGCGCC 360
XX 301 GCGCGCTATAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCATCTTTCGCGCC 360
XX
XX 361 GATGCGCATCTGATTCGCGCGCGCGCGCAAGCTTACCATGCGAGCGACCGTG 420
XX 361 GATGCGCATCTGATTCGCGCGCGCGCGCAAGCTTACCATGCGAGCGACCGTG 420
XX
XX 421 TTGCGCGAGGAGAGCGAGCGCATCTGCGCGTGAAGATACCGGCATTCGGCGCTCGGC 480
XX 421 TTGCGCGAGGAGAGCGAGCGCATCTGCGCGTGAAGATACCGGCATTCGGCGCTCGGC 480
XX
XX 481 GCGCTCTGTGCTGGAGCAGCATCAGCCATTTGTCGAATATCGCCATGTACGCCCGAC 540
XX 481 GCGCTCTGTGCTGGAGCAGCATCAGCCATTTGTCGAATATCGCCATGTACGCCCGAC 540
```

QY 541 GAACAGGTCCACGTCGGCTGTGGCCGAGCTTCAGCCCTTATGCGGCGATGCGCTATGCG 600  
Db 541 GAACAGGTCCACGTCGGCTGTGGCCGAGCTTCAGCCCTTATGCGGCGATGCGCTATGCG 600  
QY 601 CTCGACCGGAGGTCAATACCGCCGCAAGCCAGATCTTACGGGTGAGGGCGGCTGTCTAC 660  
Db 601 CTCGACCGGAGGTCAATACCGCCGCAAGCCAGATCTTACGGGTGAGGGCGGCTGTCTAC 660  
QY 661 GTGCTGCGCTGTGCGGAGCCGTTTCCCGGAGATGATCAAGGATTTGGTGTGATACGCC 720  
Db 661 GTGCTGCGCTGTGCGGAGCCGTTTCCCGGAGATGATCAAGGATTTGGTGTGATACGCC 720  
QY 721 GACAAGAGATGTTCTTCAAGGCGCGGGGGGTTTTCAGATTTTCGGGCCCGAGCGG 780  
Db 721 GACAAGAGATGTTCTTCAAGGCGCGGGGGGTTTTCAGATTTTCGGGCCCGAGCGG 780  
QY 781 CGCGCCCTGGCCGAGCCGCTCCCGGAGACCGAAGAGGAGCTGTGTGCGCGATATCGAC 840  
Db 781 CGCGCCCTGGCCGAGCCGCTCCCGGAGACCGAAGAGGAGCTGTGTGCGCGATATCGAC 840  
QY 841 CTCGGCATGATCGCTTGGCCAGAGCGCGCGGCGATCGCGCGGCGCATTTACGCGCC 900  
Db 841 CTCGGCATGATCGCTTGGCCAGAGCGCGCGGCGATCGCGCGGCGCATTTACGCGCC 900  
QY 901 GACGTACCGGCGCTGCTGTGATGATGATCGATCGCGGCCCAACCGCTGTCAAGCTTATGCC 960  
Db 901 GACGTACCGGCGCTGCTGTGATGATGATCGATCGCGGCCCAACCGCTGTCAAGCTTATGCC 960  
QY 961 GCATTCCAAACCGCAAAACGAGAGCAAGGCGGCGCGCGCTGCGCTGTGTGGCGGAA 1020  
Db 961 GCATTCCAAACCGCAAAACGAGAGCAAGGCGGCGCGCGCTGCGCTGTGTGGCGGAA 1020  
QY 1021 AGCGCGCGCGCGCGCGAGTAG 1041  
Db 1021 AGCGCGCGCGCGCGAGTAG 1041  
RESULT 5  
AD162515  
ID AD162515 standard; DNA, 1041 BP.  
XX  
AC AD162515;  
XX  
DT 22-APR-2004 (first entry)  
XX  
DE DNA encoding nitrilase polypeptide #192.  
XX  
KM Acorvastatin; (R)-ethyl 4-cyano-3-hydroxybutyrate;  
KM (R)-ethyl 4-cyano-3-hydroxybutyric acid; epichlorohydrin;  
KM 3-hydroxyglutaronitrile; 3-hydroxyglutaronitrile;  
KM 4-cyano-3-hydroxybutyric acid; ethyl-4-cyano-3-hydroxybutyric acid;  
KM mixed hyperlipidaemia; homozygous familial hypercholesterolaemia;  
KM antilipemic; gene; ds.  
XX  
OS Unidentified.  
XX  
PN WO2003106415-A2.  
XX  
PD 24-DEC-2003.  
XX  
PF 13-JUN-2003; 2003WO-US018840.  
XX  
PR 13-JUN-2002; 2002US-0389317P.  
PR 28-JUN-2002; 2002US-0392944P.  
XX  
PA (DIVE-) DIVERSA CORP.  
XX  
PI Burk M, Desantis G, Morgan B, Zhu Z;  
XX  
DR WPI; 2004-090821/09.  
XX  
DR P-PSDB; AD162516.  
XX

PT Preparation of atorvastatin comprises catalytic conversion of 3-  
PT hydroxyglutaronitrile by polypeptide with nitrilase activity, converting  
PT obtained 4-cyano-3-hydroxybutyric acid to ethyl-4-cyano-3-hydroxybutyric  
PT acid and forming atorvastatin.  
XX  
XX Claim 46; SEQ ID NO 383; 253bp; English.  
XX  
XX The present invention relates to a method for preparing an atorvastatin  
XX intermediate known as (R)-ethyl 4-cyano-3-hydroxybutyrate ((R)-ethyl 4-  
XX cyano-3-hydroxybutyric acid). The method comprises optionally converting  
XX epichlorohydrin or equivalent to 3-hydroxyglutaronitrile, catalytic  
XX conversion of 3-hydroxyglutaronitrile or equivalent to 4-cyano-3-  
XX hydroxybutyric acid with a polypeptide having nitrilase activity,  
XX converting 4-cyano-3-hydroxybutyric acid to ethyl-4-cyano-3-hydroxybutyric  
XX acid, and converting this to (R)-ethyl 4-cyano-3-hydroxybutyrate. The  
XX method involves whole cell processes, cell lysate process, "one pot"  
XX processes, and "multi-pot" processes using a variety of parameters.  
XX Acorvastatin is used, in conjunction with dietary restriction, in the  
XX management of hyperlipidaemia, including hypercholesterolaemia, mixed  
XX dyslipidaemia and homozygous familial hypercholesterolaemia. The present  
XX sequence encodes a nitrilase polypeptide obtained from an environmental  
XX sample.  
SQ Sequence 1041 BP; 172 A; 340 C; 357 G; 172 T; 0 U; 0 Other;  
Query Match 100.0%; Score 1041; DB 12; Length 1041;  
Best Local Similarity 100.0%; Pred. No. 1, 1e-197;  
Matches 1041; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ATGTGAGAGCCATGACGAAGTATCGCGCGCGCGGTGACGCGCGCGGTTCCTC 60  
Db 1 ATGTGAGAGCCATGACGAAGTATCGCGCGCGCGGTGACGCGCGCGGTTCCTC 60  
QY 61 GATTCGACCCGACAGTCCGAGAAAGGATGCGCTGATGAGAGCGCGCGCAACGAGAC 120  
Db 61 GATTCGACCCGACAGTCCGAGAAAGGATGCGCTGATGAGAGCGCGCGCAACGAGAC 120  
QY 121 GTGCGCTGATGCGATTCGCCAGAGACTTGGATTCCCGGCTATTCCTTTGGATTTGGCTG 180  
Db 121 GTGCGCTGATGCGATTCGCCAGAGACTTGGATTCCCGGCTATTCCTTTGGATTTGGCTG 180  
QY 181 GCGCGCCGCGCTTGGGAGCATGCGCTTGTGTCAGCGCTATTTTCGAGATTCGTCGCGC 240  
Db 181 GCGCGCCGCGCTTGGGAGCATGCGCTTGTGTCAGCGCTATTTTCGAGATTCGTCGCGC 240  
QY 241 GCGAGCAAGCATGTCGAGGCGCTTGGGATGCGCGCGCGCGCAAGCGCATGTCGTG 300  
Db 241 GCGAGCAAGCATGTCGAGGCGCTTGGGATGCGCGCGCGCGCAAGCGCATGTCGTG 300  
QY 301 GCGGCTATAGCAGAGCGCGCGCGCGCGCGCGCTTATATGAGCGAGCGATCTTCGCGCC 360  
Db 301 GCGGCTATAGCAGAGCGCGCGCGCGCGCGCGCTTATATGAGCGAGCGATCTTCGCGCC 360  
QY 361 GATGCGATCTGATGCGCGCGCGCGCGCGCGCGCTTATATGAGCGAGCGATCTTCGCGCC 420  
Db 361 GATGCGATCTGATGCGCGCGCGCGCGCGCGCGCTTATATGAGCGAGCGATCTTCGCGCC 420  
QY 421 TTCGCGAGGAGACGCGAGCGCATCTTCGCGGTGACAGATACGCGCATTCGCGCGCTTCGCGC 480  
Db 421 TTCGCGAGGAGACGCGAGCGCATCTTCGCGGTGACAGATACGCGCATTCGCGCGCTTCGCGC 480  
QY 481 GCGCTCTGTTGCTGGAGGACATCCAGCCATTTTCGAAATACGCGCATGTCGCGCGGAC 540  
Db 481 GCGCTCTGTTGCTGGAGGACATCCAGCCATTTTCGAAATACGCGCATGTCGCGCGGAC 540  
QY 541 GAACAGGTCCACGTCGGCTGTGGCCGAGCTTCAGCCCTTATGCGGCGATGCGCTATGCG 600  
Db 541 GAACAGGTCCACGTCGGCTGTGGCCGAGCTTCAGCCCTTATGCGGCGATGCGCTATGCG 600  
QY 601 CTCGACCGGAGGTCAATACCGCCGCAAGCCAGATCTTACGGGTGAGGGCGGCTGTCTAC 660  
Db 601 CTCGACCGGAGGTCAATACCGCCGCAAGCCAGATCTTACGGGTGAGGGCGGCTGTCTAC 660

```
QY 661 GTGCTGGCTGTCGTGCGGACCGGTTTCGCGGAGATGATCAAGATTGTTGATACGCC 720
D 661 GTGCTGGCTGTCGTGCGGACCGGTTTCGCGGAGATGATCAAGATTGTTGATACGCC 720
QY 721 GACAAGAGATGTTCTCTCAAGGCGGCGGCGGTTTTCGATGATTTTCGCGCCGAGCGG 780
D 721 GACAAGAGATGTTCTCTCAAGGCGGCGGCGGTTTTCGATGATTTTCGCGCCGAGCGG 780
QY 781 CGCGCCCTGCGCGGACCGGCTCTCCGGAACCGAAGAGGACTGTGTGCGCGATATGAC 840
D 781 CGCGCCCTGCGCGGACCGGCTCTCCGGAACCGAAGAGGACTGTGTGCGCGATATGAC 840
QY 841 CTGCGCATGATCGGCTTGGCGCAAGCGGCGGCGGATCGCGGCGGCGCACTATTACCGGCC 900
D 841 CTGCGCATGATCGGCTTGGCGCAAGCGGCGGCGGATCGCGGCGGCGCACTATTACCGGCC 900
QY 901 GACGTAAACGCGGCTGCTGCTGATGACGTCGCGGCCAACCGCGTGTCAAGCTTGATGCC 960
D 901 GACGTAAACGCGGCTGCTGCTGATGACGTCGCGGCCAACCGCGTGTCAAGCTTGATGCC 960
QY 961 GCATTGCAACCGCAAAACGAGAACAGGCGGACCGCGCGCTGCGCGGCGGCGAA 1020
D 961 GCATTGCAACCGCAAAACGAGAACAGGCGGACCGCGCGCTGCGCGGCGGCGAA 1020
QY 1021 AGCGCCGCGCGCGCGCGCAGTAG 1041
D 1021 AGCGCCGCGCGCGCGCGCAGTAG 1041
```

RESULT 6  
AD164636  
ID AD164636 standard; DNA; 1041 BP.

XX AD164636:

DT 22-APR-2004 (first entry)

DE DNA encoding nitrilase seq id 192.

XX (R)-ethyl 4-cyano-3-hydroxybutyric acid; nitrile hydrolysis;  
KM carboxylic acid; cyanohydrin moiety hydrolysis;  
KM ammonitrile moiety hydrolysis; chiral alpha-hydroxy acid molecule;  
KM chiral amino acid molecule; (S)-ethyl 4-cyano-3-hydroxybutyric acid;  
KM (R)-mandelic acid; (S)-mandelic acid; (S)-phenyl lactic acid derivative;  
KM (R)-phenyl lactic acid derivative; & enantiomeric excess;  
KM & diastereomeric excess; food additive; drug intermediate; ds; nitrilase;  
KM gene.

XX Unidentified.

OS US2004014195-A1.

XX 22-JAN-2004.

PD 15-MAY-2003; 2003US-00440523.

XX 29-DEC-1999; 99US-0173609P.  
PR 07-DEC-2000; 2000US-0254414P.  
PR 28-DEC-2000; 2000US-00751299.  
PR 21-JUN-2001; 2001US-0300189P.  
PR 30-JUL-2001; 2001US-0309006P.  
PR 22-JAN-2002; 2002US-0351136P.  
PR 15-MAY-2002; 2002US-00146772.  
PR 09-SEP-2002; 2002US-00241742.

XX (DIVE-) DIVERSA CORP.

XX Desantis G, Short JM, Burk MJ, Wong K, Farwell R, Chatman K;

XX WPI; 2004-121569/12.

DR P-PSDB; AD164637.

XX Novel isolated or recombinant polypeptide having nitrilase activity,

PT useful in production of food additives.  
XX Claim 1; SEQ ID NO 383; 105pp; English.  
PS  
XX The invention describes an isolated or recombinant polypeptide (1)  
CC comprising amino acid having a sequence at least 50 % identical to a  
CC sequence (S1) available in electronic form (EC) from the following web  
CC site ftp.biodata.uspto.gov/sequence.html?docid=2004014195, or its  
CC variants, having one or more mutations at residue 55 lys, Gly or Glu, at  
CC residue 60 glutamic acid, at residue 111 ser, their combinations or  
CC fragments. (1) is useful for: producing an (R)-ethyl 4-cyano-3-  
CC hydroxybutyric acid; hydrolysing a nitrile to a carboxylic acid;  
CC hydrolysing cyanohydrin moiety or an ammonitrile moiety; producing a  
CC chiral alpha-hydroxy acid molecule or a chiral amino acid molecule;  
CC producing an (S)-ethyl 4-cyano-3-hydroxybutyric acid; producing an (R)-  
CC mandelic acid or (S)-mandelic acid; producing (S)-phenyl lactic acid  
CC derivative or an (R)-phenyl lactic acid derivative; modifying a molecule;  
CC and for identifying a modified compound. The inventive method is useful  
CC for monitoring or determining & enantiomeric excess or & diastereomeric  
CC excess. (1) is useful in the production of food additives and drug  
CC intermediates. This sequence encodes a nitrilase of the invention.  
XX  
SQ Sequence 1041 BP; 172 A; 340 C; 357 G; 172 T; 0 U; 0 Other;

Query Match 100.0%; Score 1041; DB 12; Length 1041;  
Best Local Similarity 100.0%; Pred. No. 1,1e-197;  
Matches 1041; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 ATGTCGAGCCCATGACGAAGTATCGGCGCGGCGGTGACAGCCCGCGGTGTTCTC 60
D 1 ATGTCGAGCCCATGACGAAGTATCGGCGCGGCGGTGACAGCCCGCGGTGTTCTC 60
QY 61 GATCTGACCGGACAGTGCAGAAAGGATGCGGCTGATGACGAGCGGCGCAAGCGAGC 120
D 61 GATCTGACCGGACAGTGCAGAAAGGATGCGGCTGATGACGAGCGGCGCAAGCGAGC 120
QY 121 GTGCGCTGATGATGATCCCAAGACTTGGATTTCCGCTATTGATGATGCTG 180
D 121 GTGCGCTGATGATGATCCCAAGACTTGGATTTCCGCTATTGATGATGCTG 180
QY 181 GGGGCGCGGCTTTGGGGCATGCGCTTCGTCAGCGCTATTTGAGAAATTCGTCGCGC 240
D 181 GGGGCGCGGCTTTGGGGCATGCGCTTCGTCAGCGCTATTTGAGAAATTCGTCGCGC 240
QY 241 GCGAGCAACAGTGGGAGGCGCTTGGCGATGCGGCGCGCGCGCAAGCGATGATGCTG 300
D 241 GCGAGCAACAGTGGGAGGCGCTTGGCGATGCGGCGCGCGCGCAAGCGATGATGCTG 300
QY 301 GCGGCTATGAGGAGCGCGCGGCGGAGCTCTATATGAGGCGCAAGCGATGCTTGGCGCC 360
D 301 GCGGCTATGAGGAGCGCGCGGCGGAGCTCTATATGAGGCGCAAGCGATGCTTGGCGCC 360
QY 361 GATGGGATCTGATGCGCGCGGCGGCGCAAGCTCAAGCTTCAAGCGAGCGAGCGGCGG 420
D 361 GATGGGATCTGATGCGCGCGGCGGCGCAAGCTTCAAGCTTCAAGCGAGCGAGCGGCGG 420
QY 421 TTCGCGGAGGAGAGCGGCGGCGCATCTCGGAGTGAAGATCCGATCGGCGGCTCGGC 480
D 421 TTCGCGGAGGAGAGCGGCGGCGCATCTCGGAGTGAAGATCCGATCGGCGGCTCGGC 480
QY 481 GGGCTCTGTTGCTGGAGAGACATCCAGACATTTGAAATAGCCATGATACCGCGCGAC 540
D 481 GGGCTCTGTTGCTGGAGAGACATCCAGACATTTGAAATAGCCATGATACCGCGCGAC 540
QY 541 GAAACAGTTCACGTGCGGCTGAGCGGAGCTTCAAGCTTATCGGCGGCGATGAGCG 600
D 541 GAAACAGTTCACGTGCGGCTGAGCGGAGCTTCAAGCTTATCGGCGGCGATGAGCG 600
QY 601 CTCGAGACGAGGATCAATACCGCGCAAGCGATCTACGCGGTGAGAGGCGGCTGTAAC 660
D 601 CTCGAGACGAGGATCAATACCGCGCAAGCGATCTACGCGGTGAGAGGCGGCTGTAAC 660
QY 661 GTGCTGGGCTGCTGGCGGACCGTTTCGCGGAGATGATCAAGTATTTGGTGAATACGCC 720
```



Db	661	GTGCTGGGGTGTGTGGCCGACCGTTTTCGCTCGGAGATGATCAAGGATATTTGTGTGATACGGCC	720
QY	721	GACAAAGAGATGTTTCTCAAGGCGGGCGGGTTTGTGCATGATTTTGGGCGCCGACGGC	780
Db	721	GACAAAGGATGTTTCTCAAGGCGGGCGGGTTTGTGCATGATTTTGGGCGCGACGGC	780
QY	781	CGCGCCCTGGGCCGAGCCGCTCCCGGAGACCGAAGAGGGACTGTGTCGCCGATATCGAC	840
Db	781	CGCGCCCTGGGCCGAGCCGCTCCCGGAGACCGAAGAGGGACTGTGTCGCCGATATCGAC	840
QY	841	CTCGGCATGATCGCGTTGGCCAAAGCGGGCGGCGCATCCGGGCGGGCACTATTTCACGGCCC	900
Db	841	CTCGGCATGATCGCGTTGGCCAAAGCGGGCGGCGCATCCGGGCGGGCACTATTTCACGGCCC	900
QY	901	GACGTAACGGCGGCTGTCTGTGANTGACGTCGCGGCCAACGCGTCGTGACGCTTGATGCC	960
Db	901	GACGTAACGGCGGCTGTCTGTGANTGACGTCGCGGCCAACGCGTCGTGACGCTTGATGCC	960
QY	961	GCATTTCGAACCGGCCAAACGAGGACAAAGGGCGACGCGCCCGCTGCGCGTGTGGCGGAA	1020
Db	961	GCATTTCGAACCGGCCAAACGAGGACAAAGGGCGACGCGCCCGCTGCGCGTGTGGCGGAA	1020
QY	1021	AGCGCGCGCGCGCGCAGTAG	1041
Db	1021	AGCGCGCGCGCGCGCAGTAG	1041

ammonia. Nitrilases have commercial utility as biocatalysts for use in the synthesis of enantiomerically pure aromatic and aliphatic amino acids, as well as hydroxy acids, which are important for the development of chiral medicines. Furthermore, the present invention describes nitrilases, isolated from mesophilic microorganisms, that have improved activity and stability at increased pH and temperature. They are also inexpensive, efficient catalysts, have broad substrate specificity and are capable of chiral differentiation. This polynucleotide is a DNA sequence that encodes a nitrilase enzyme of the invention.



KW nitrilase; nitrile; carboxylic acid; chemical process; pH; temperature;  
KM enantioselective transformation; gene; ds.  
XX Unidentified.  
XX  
XX WO2003097810-A2.  
XX  
XX 27-NOV-2003.  
XX  
XX 15-MAY-2003; 2003WO-US015712.  
XX PF  
XX 15-MAY-2002; 2002US-00146772.  
PR 09-SEP-2002; 2002US-00241742.  
XX  
XX (DIVE-) DIVERSA CORP.  
PA  
PI Desantis G, Short JM, Burk M, Wong K, Farwell R, Chatman K;  
XX MPI: 2004-090637/09.  
DR P-PSDB; ADG933569.  
XX  
XX New isolated or recombinant nucleic acid encoding a polypeptide having  
PT nitrilase activity, useful for screening enantioselective transformation.  
XX  
XX Claim 44; SEQ ID NO 33; 295bp; English.  
PS  
CC This invention is related to a novel isolated or recombinant nucleic acid  
CC encoding a protein having nitrilase activity. Nitrilase's are capable of  
CC converting nitrile's directly to carboxylic acids and have great  
CC potential for use in industrial chemical processes. The isolated  
CC nitrilase proteins of the invention have increased activity and stability  
CC at increased pH and temperature when compared to those conventionally  
CC used. In addition, the nucleic acid of the invention is useful for  
CC screening enantioselective transformation. The present sequence is that  
CC of a DNA sequence which encodes a nitrilase enzyme of the invention.  
XX  
SQ Sequence 1026 BP; 158 A; 347 C; 357 G; 164 T; 0 U; 0 Other;  
  
Query Match 65.5%; Score 682; DB 12; Length 1026;  
Best Local Similarity 79.4%; Pred. No. 2.5e-126;  
Matches 808; Conservative 0; Mismatches 210; Indels 0; Gaps 0;  
  
QY 1 ATGTCGAGAGCCGACGAGAAAGCATGCGCGCGGCGGCGCGCGCGCGCTTCTC 60  
DB 1 ATGTTAACTCCCGTAGCCAGATATGCGCGCGGCGGCGCGCGCGCATCTTTTCTC 60  
  
QY 61 GATCTGACCGCAGCATGCGAGAAAGCATGCGCGCGCTGATCGAGCAGCGCGCAAGCAGGAC 120  
DB 61 GATCTGACCGCAGCATGCGAGAAAGCATGCGCGCGCTGATCGAGCAGCGCGCGCATGAT 120  
  
QY 121 GTGCGCCTGATCGCATTTCCAGAGAATTGATTTCCCGGCTATCCCTTTTGGATATGAGCTG 180  
DB 121 GTGCGCCTGATCGCGCTTTCCGGAACCTGATTTCCCGGCTATCCGCTCGGATCTGGCTC 180  
  
QY 181 GCGCGCGCGCGCTTGGGCGATGCGCTTCTGTCAGCGCGCTATTTGAGAAATTCGCTGCGCC 240  
DB 181 GCGCTGCGCGCGCTTGGGCGATGCGCTTCTGTCAGCGCGCTATTTGAGAAATTCGCTGCGCC 240  
  
QY 241 GGCAGCAGACGATGCGCAGCGCGCTGCGGATGCGCGCGCGCGCGCGCGCATGATGCTGCG 300  
DB 241 GGCAGCAGACGATGCGCAGCGCGCTGCGGATGCGCGCGCGCGCGCGCGCATGATGCTGCG 300  
  
QY 301 GCCCGCTATAGGAGCGCGCGCGCGCGCGCGCGCTTATATGCGCGCAGCGCATCTTCCGCCCC 360  
DB 301 GTCGCGCTTACGAGGAGCGCGCGCGCGCGCGCGCGCTTATATGCGCGCAGCGCATCTTCCGCCCC 360  
  
QY 361 GATGCGCATCTGATGCG 420  
DB 361 GATGCGCATCTGATGCG 420  
  
QY 421 TTCCGCGAGGAGGAGCAGCG 480  
DB 421 TTCCGCGAGGAGGAGCAGCG 480

QY 481 GCGCTCTGTTGCTGGAGACATCCAGCATTTGCAAAATACCGCATGTACCGCGCGAC 540  
DB 481 GCGCTCTGCTGCTGGAGACATCCAGCATTTGCAAAATACCGCATGTATGCGCGCGAC 540  
  
QY 541 GAACAGGTCCAGCTGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 600  
DB 541 GAACAGGTCCAGCTGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 600  
  
QY 601 CTCGAGCCGAGAGTCAATACCGCGCGCAAGCCAGATCTACGCGGTGAGAGCGCGCTGTAC 660  
DB 601 CTCGAGCCGAGAGTCAATACCGCGCGCAAGCCAGATCTACGCGGTGAGAGCGCGCTGTAC 660  
  
QY 661 GTGCTGCGGTGCG 720  
DB 661 GTGCTGCGGTGCG 720  
  
QY 721 GACAAGAGATGTTCTCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 780  
DB 721 GACAAGAGATGTTCTCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 780  
  
QY 781 CGCGCGCTGCG 840  
DB 781 CGCGCGCTGCG 840  
  
QY 841 CTCGCGCATGATGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 900  
DB 841 CTCGCGCATGATGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 900  
  
QY 901 GACGTTAAGCGCGGTGCTGCTGATGAGCGTCCGCGCGCGCGCGCGCGCGCGCGCGCG 960  
DB 901 GACGTTAAGCGCGGTGCTGCTGATGAGCGTCCGCGCGCGCGCGCGCGCGCGCGCGCG 960  
  
QY 961 GCATTTGAAACCGCAAAAGAGACAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1018  
DB 961 GCATTTGAAACCGCAAAAGAGACAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1018  
  
RESULT 10  
AD162165  
ID AD162165 standard; DNA; 1026 BP.  
XX  
XX AC AD162165;  
XX  
XX DT 22-APR-2004 (first entry)  
XX  
XX DE DNA encoding nitrilase polypeptide #17.  
XX  
XX KM Acetovastatin; (R)-ethyl 4-cyano-3-hydroxybutyrate;  
XX (R)-ethyl 4-cyano-3-hydroxybutyric acid; epichlorohydrin;  
XX 3-hydroxyglutaronitrile; 3-hydroxyglutaronitrile;  
XX 4-cyano-3-hydroxybutyric acid; ethyl-4-cyano-3-hydroxybutyric acid;  
XX mixed hyperlipidaemia; homozygous familial hypercholesterolaemia;  
XX antiliphaemic; gene; ds.  
OS Unidentified.  
XX  
XX PN WO2003106415-A2.  
XX  
XX PD 24-DEC-2003.  
XX  
XX PF 13-JUN-2003; 2003WO-US018840.  
XX  
XX PR 13-JUN-2002; 2002US-0389317P.  
XX  
XX PR 28-JUN-2002; 2002US-0392944P.  
XX  
XX PA (DIVE-) DIVERSA CORP.  
XX  
XX PI Burk M, Desantis G, Morgan B, Zhu Z;  
XX MPI: 2004-090821/09.  
XX  
XX DR P-PSDB; AD162166.  
XX

PT Preparation of atorvastatin comprises catalytic conversion of 3-hydroxyglutaronitrile by polypeptide with nitrilase activity, converting 4-cyano-3-hydroxybutyric acid to ethyl-4-cyano-3-hydroxybutyric acid and forming atorvastatin.

XX Claim 46; SEQ ID NO 33; 253bp; English.

CC The present invention relates to a method for preparing an atorvastatin intermediate known as (R)-ethyl 4-cyano-3-hydroxybutyrate ((R)-ethyl 4-cyano-3-hydroxybutyric acid). The method comprises optionally converting epichlorohydrin or equivalent to 3-hydroxyglutaronitrile, catalytic conversion of 3-hydroxyglutaronitrile or equivalent to 4-cyano-3-hydroxybutyric acid with a polypeptide having nitrilase activity, converting 4-cyano-3-hydroxybutyric to ethyl-4-cyano-3-hydroxybutyric acid, and converting this to (R)-ethyl 4-cyano-3-hydroxybutyrate. The method involves whole cell processes, cell lysate processes, "one pot" processes, and "multi-pot" processes using a variety of parameters. Atorvastatin is used, in conjunction with dietary restriction, in the management of hyperlipidaemia, including hypercholesterolaemia, mixed dyslipidaemia and homozygous familial hypercholesterolaemia. The present sequence encodes a nitrilase polypeptide obtained from an environmental sample.

CC Sequence 1026 BP; 158 A; 347 C; 357 G; 164 T; 0 U; 0 Other;

Query Match 65.5%; Score 682; DB 12; Length 1026;

Best Local Similarity 79.4%; Pred. No. 2.5e-126; Matches 808; Conservative 0; Mismatches 210; Indels 0; Gaps 0;

QY 1 ATGTCGAGAGCCATGAGAGATATCGGCGCGCGGTCGACGCGCGCGGTTCCTC 60  
 DB 1 ATGTTAAGTCCCGAGAGATATCGGCGCGCGCGGTCGACGCGCGCGCATCTTTCTC 60  
 QY 61 GATTCGACGCGACGTCGAGAAAGCATGCGCTGATTCGAGAGCGCGCGCAAGCAGAC 120  
 DB 61 GATTCGACGCGACGTCGAGAAAGCATGCGCTGATTCGAGAGCGCGCGCGCAGACGAT 120  
 QY 121 GTGCGCTGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTC 180  
 DB 121 GTGCGCTGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTC 180  
 QY 181 GCGCGCGCGCGCTTGGGCGATGCGCTTGGGCGATGCGCTTGGGCGATGCGCTTGGGCG 240  
 DB 181 GCGCGCGCGCGCTTGGGCGATGCGCTTGGGCGATGCGCTTGGGCGATGCGCTTGGGCG 240  
 QY 241 GCGCGCGCGCGCTTGGGCGATGCGCTTGGGCGATGCGCTTGGGCGATGCGCTTGGGCG 300  
 DB 241 GCGCGCGCGCGCTTGGGCGATGCGCTTGGGCGATGCGCTTGGGCGATGCGCTTGGGCG 300  
 QY 301 GCGCGCGCGCGCTTGGGCGATGCGCTTGGGCGATGCGCTTGGGCGATGCGCTTGGGCG 360  
 DB 301 GCGCGCGCGCGCTTGGGCGATGCGCTTGGGCGATGCGCTTGGGCGATGCGCTTGGGCG 360  
 QY 361 GATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTC 420  
 DB 361 GATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTCGATTC 420  
 QY 421 TTGCGCGAGGAGCG 480  
 DB 421 TTGCGCGAGGAGCG 480  
 QY 481 GCGCTCTGTTGCTGGAGACATTCGAGCATTTGCAATATGCGCATGTATGCGCGCGAC 540  
 DB 481 GCGCTCTGTTGCTGGAGACATTCGAGCATTTGCAATATGCGCATGTATGCGCGCGAC 540  
 QY 541 GAACAGGTCACGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 600  
 DB 541 GAACAGGTCACGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 600  
 QY 601 CTGCGACCGGAGGTCATATCCGCGCAGCAGATCTACCGGTCGAGGCGGCTCTAC 660  
 DB 601 CTGCGACCGGAGGTCATATCCGCGCAGCAGATCTACCGGTCGAGGCGGCTCTAC 660

QY 661 GTGCTGCGCTGTCGTCGCGGACCGTTTCCCGAGATGATCAAGGATATGTCGATACGCC 720  
 DB 661 GTGCTGCGCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 720  
 QY 721 GACAGAGATGTTCTCAAGCG 780  
 DB 721 GACAGAGATGTTCTCAAGCG 780  
 QY 781 CGCGCGCTGCG 840  
 DB 781 CGCGCGCTGCG 840  
 QY 841 CTGCGCATATTCGCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 900  
 DB 841 CTGCGCATATTCGCTGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCGTCG 900  
 QY 901 GACGTACCG 960  
 DB 901 GACGTACCG 960  
 QY 961 GCATTCGACCGCAAAACGAGACAAAGGCGACGCGCGCGCGCGCGCGCGCGCGCGCG 1018  
 DB 961 GCATTCGACCGCGCAAAACGAGACAAAGGCGACGCGCGCGCGCGCGCGCGCGCGCGCG 1018

RESULT 11

AD164286 standard; DNA; 1026 BP.

AD164286;

22-APR-2004 (first entry)

DNA encoding nitrilase seq id 17.

(R)-ethyl 4-cyano-3-hydroxybutyric acid; nitrile hydrolysis;  
 carboxylic acid; cyanohydrin moiety hydrolysis;  
 amino nitrile moiety hydrolysis; chiral alpha-hydroxy acid molecule;  
 chiral amino acid molecule; (S)-ethyl 4-cyano-3-hydroxybutyric acid;  
 (R)-mandelic acid; (S)-mandelic acid; (S)-phenyl lactic acid derivative;  
 (R)-phenyl lactic acid derivative; % enantiomeric excess;  
 % diastereomeric excess; food additive; drug intermediate; ds; nitrilase;  
 gene.

Unidentified.

US2004014195-A1.

22-JAN-2004.

15-MAY-2003; 2003US-00440523.

29-DEC-1999; 99US-0173609P.

07-DEC-2000; 2000US-0254414P.

28-DEC-2000; 2000US-0075129P.

21-JUN-2001; 2001US-0300189P.

30-JUL-2001; 2001US-030906P.

22-JAN-2002; 2002US-0351336P.

15-MAY-2002; 2002US-0014677P.

09-SEP-2002; 2002US-00241742.

(DIVE-) DIVERSA CORP.

Desantle G, Short JM, Burk MJ, Wong K, Farwell R, Chatman K;  
 WPI, 2004-121569/12.  
 P-PSDB; AD164287.  
 Novel isolated or recombinant polypeptide having nitrilase activity,  
 useful in production of food additives.  
 Claim 1; SEQ ID NO 33; 105bp; English.

CC The invention describes an isolated or recombinant polypeptide (I) comprising amino acids having a sequence at least 50 % identical to a sequence (S1) available in electronic form (EC) from the following web site [ftp.segdate.uspto.gov/sequence.html?DocID=2004014195](http://ftp.segdate.uspto.gov/sequence.html?DocID=2004014195), or its variants, having one or more mutations at residue 55 Lys, Gly or Glu, at residue 60 glutamic acid, at residue 111 Ser, their combinations or fragments. (I) is useful for: producing an (R)-ethyl 4-cyano-3-hydroxybutyric acid; hydrolysing a nitrile to a carboxylic acid; hydrolysing cyanohydrin moiety or an aminonitrile moiety; producing a chiral alpha-hydroxy acid molecule or a chiral amino acid molecule; producing an (S)-ethyl 4-cyano-3-hydroxybutyric acid; producing an (R)-mandelic acid or (S)-mandelic acid; producing (S)-phenyl lactic acid derivative or an (R)-phenyl lactic acid derivative; modifying a molecule; and for identifying a modified compound. The inventive method is useful for monitoring or determining % enantiomeric excess or % diastereomeric excess. (I) is useful in the production of food additives and drug intermediates. This sequence encodes a nitrilase of the invention.

CC Sequence 1026 BP; 158 A; 347 C; 357 G; 164 T; 0 U; 0 Other;

Query Match 65.5%; Score 682; DB 12; Length 1026;  
Best Local Similarity 79.4%; Pred. No. 2.5e-126;  
Matches 808; Conservative 0; Mismatches 210; Indels 0; Gaps 0;

```

QY 1 ATGTGGAGGCGCATGAGAGTATCGGGGCGGGTGCAGGCGCGGCGGTTCCTC 60
DB 1 ATGTAAATCGCCGAGACAGTATCGCGCGCGGGTGCAGGCGCGCATCTTTCTC 60
QY 61 GATCTGCACCGCAGTCTGAGAAAGCGATCGGCTGATCGAGAGCGGCGCAAGAGAC 120
DB 61 GATCTGACCGGACCGTCTGAGAGAGCATGCGCATCTCGAGAGCGGCGCGAGAGAT 120
QY 121 GTGCGCTGATCGCATTCGAGAGACTTGAGATTCGCGGCTTATCCCTTTGATATG 180
DB 121 GTGCGCTGATCGCATTCGCGGAAACCTGGAATTCGCGCTGATCTGATCTGGCTC 180
QY 181 GGGGCGCGCGCTTGGGGCATGCGCTTGTCTGACAGGCTATTCAGAGATTCGCTG 240
DB 181 GGGTGGCGCGCTTGGGGCATGCGCTTGTCTGACAGGCTATTCAGAGATTCGCTG 240
QY 241 GGGAGAGAGAGTGGAGAGCGCTTGGCGAGATGCGGCGCGCGCGCATGATGTCGT 300
DB 241 GGGAGAGAGAGTGGAGAGCGGATGCGGATGCGGCGCGCGCGCATGATGTCGT 300
QY 301 GCGGCGCTATAGCGAGCGCGCGCGCGAGCCTTATATGAGCGCATCTTCGCGCC 360
DB 301 GTCGGCTTACGAGCGCGCGGAGGAGCCTTACATGAGCGAGGATCTTCGCGCC 360
QY 361 GATGGGATTCGATCGCGCGCGCGCGAGCTCAAGCTTACATGCGGAGCGCACCGTG 420
DB 361 GATGGGATTCGATCGCGCGCGCGCGAGCTCAAGCTTACATGCGGAGCGCACCGTG 420
QY 421 TTGCGGAGGAGAGCGGAGCGCATCTGCGGATGAGATCCGCGCATCGGCGCTGCG 480
DB 421 TTGCGGAGGAGAGCGGAGCGCATCTGCGGATGAGATCCGCGCATCGGCGCTGCG 480
QY 481 GCGCTCTGTTGTGGAGCATTCAGACCTTGTGAAATATCGGCATGATCGCGCGGAC 540
DB 481 GCGCTCTGTTGTGGAGCATTCAGACCGCTCTCGAAATATCGGCATGATGCGCGGAC 540
QY 541 GAACAGGTCCAGTCCGCTGTGCGCGAGCTTCAAGCTTATGCGGCGATGCGCTATGCG 600
DB 541 GAACAGGTCCAGTCCGCTGTGCGCGAGCTTCAAGCTTATGCGGCGATGCGCTATGCG 600
QY 601 CTGCGACCGAGAGTCAATACCGCGCGAGCATGATCAAGCGGTCGAGCGGCTGTAC 660
DB 601 CTGCGCGCGAGAGTGAACACCGCGCGAGCATGATCAAGCGGTCGAGCGGCTGTAC 660
QY 661 GTCTGGCGCTGTGCGCGGAGCGGTTTTCGCGAGATGATCAAGGATTTGATGATCGCCC 720
DB 661 GTCTGGCGCTGTGCTGTGCGAGACCGGAGATCTGAAGGATGATGATGATGATGATG 720
QY 721 GACAGAGAGATGTTCTCAAGGCGCGGCGGTTTTCGCGATGATTTTCGCGCGCGAGCGC 780

```

```

DB 721 GACAGAGAGCGTGTGCTGCTCGCGCGGGGGTTCTCATGATCTTCGCGCGAGCGC 780
QY 781 CGGCGCTTGGCGAGCGCGCTCCGAGACCGAAGAGAGTCTGTGCTGCGATATGAC 840
DB 781 CGGCGCTTGGCGAGCGCGCTCCGAGACCGAAGAGAGTCTGTGCTGCGATATGAC 840
QY 841 CTGCGCATGATCGCGTGTGGCGCAAGGCGGCGCGATCCGCGGCGCATATTCAGCGGCC 900
DB 841 CTGCGCGCATGATCGCGTGTGGCGCAAGGCGGCGCGATCCGCGGCGCATATTCAGCGGCC 900
QY 901 GACGTAACGCGGCTGTGCTGATGATGATCGGCGCGCAAGCGGCTGTGATGATGCGC 960
DB 901 GACGTAACGCGGCTGTGCTGATGATGATCGGCGCGCAAGCGGCTGTGATGATGCGC 960
QY 961 GCATTGGAACCGGAAACGAGAGCAAGAGGAGCGCGCGCGCTGCGCGTGTGCGCG 1018
DB 961 CGCTCGAGGTGTGCGAGAGCGAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCG 1018

```

RESULT 12  
ADCC3944  
ID ADCC3944 standard; DNA; 1062 BP.

AC ADCC3944;  
XX  
DT 18-DEC-2003 (first entry)  
XX  
DE DNA sequence (SeqID 211) encoding a nitrilase enzyme.

KW gene; de; nitrilase; nitrile; cyanohydrin; ammonia; biocatalyst;  
KW enantiomer; chiral medicine.

OS Unidentified.

PN NC003000840-A2.

XX  
PD 03-JUN-2003.

XX  
PF 15-MAY-2002; 2002MO-US015983.

XX  
PR 21-JUN-2001; 2001US-0300189P.

XX  
PR 30-JUL-2001; 2001US-0309006P.

XX  
PR 22-JUN-2002; 2002US-0351336P.

XX  
PA (DIVE-) DIVERSA CORP.

XX  
PA (MADP/) MADDEN D.

XX  
PI Madden M, Desantis G, Chaplin JA, Weiner DP, Milan A, Chi E;

XX  
PI Short JM, Burk M;

XX  
DR WPI; 2003-201417/19.

XX  
DR P-PSDB; ADCC3945.

XX  
PT Novel nitrilase polypeptide, useful for making (R)- or (S)-ethyl-4-cyano-3-hydroxybutyric acid or (R)- or (S)-mandelic acid or (S)- or (R)-phenyl lactic acid derivative and for producing pharmaceutical composition, and food additive.

XX  
PS Claim 1; SEQ ID NO 211; 560pp; English.

XX This invention relates to nitrilases and the nucleic acids that encode these enzymes thereof. Specifically, it refers to polypeptides that exhibit nitrilase activity, i.e. the ability to directly hydrolyse nitriles or cyanohydrins into their corresponding carboxylic acids and ammonia. Nitrilases have commercial utility as biocatalysts for use in the synthesis of enantiomerically pure aromatic and aliphatic amino acids, as well as hydroxy acids, which are important for the development of chiral medicines. Furthermore, the present invention describes nitrilases, isolated from mesophilic microorganisms, that have improved activity and stability at increased pH and temperature. They are also inexpensive, efficient catalysts, have broad substrate specificity and are capable of chiral differentiation. This polynucleotide is a DNA

CC sequence that encodes a nitrilase enzyme of the invention.

XX Sequence 1062 BP; 186 A; 355 C; 341 G; 180 T; 0 U; 0 Other;

Query Match 46.7%; Score 486; DB 10; Length 1062;

Best Local Similarity 67.8%; Pred. No. 2,3e-87;

Matches 697; Conservative 0; Mismatches 325; Indels 6; Gaps 1;

```
QY 17 CGAAGTATCCGCGCGCGCGGTGACGCCCGCGCGGTCTCTGATCTGACCGCACAG 76
DB 17 CGAATATCAAAAGTGGCCGCGCGGTGACGCCCGCGCGGTCTCTGACCGTGGACCGCTCGG 76
QY 77 TCGAAGAAAGCGATCGGCTTGATCGAGCAGCGCGCAAGCAGAGAGTGGCTGATCGCAT 136
DB 77 TCGAAGAAAGCGCGCTTTCATCGACGAGCGCGCGCGCGCGCGCTTATCGCT 136
QY 137 TCCCAAGATTTGATTTCCCGGCTATCCCTTTTGGATATGCTGGCGCGCGCTTGGG 196
DB 137 TCCCGAGACCTGATACCCGGTTACCCCTGGTGGATCTGATAGGCGCGCGCGCTGGG 196
QY 197 GCAT-----GCGCTTCTGTCAGCGCTATTTCGAAATTCGCTGTCGCGGCGCAAGC 250
DB 197 CTATCATGCGCGGCTTGTCTCGCGCTATTTCGACAACTGCTCAGCTACGACAGCCCGC 256
QY 251 AGTGGCAGAGCGCTGCGCGATGCGCGCGCGCGCGCGCGCGCGCATGTGTGGCGCTATA 310
DB 257 AGGCGGAAAGCTTCGCGCGCGCGCGCAAGCGCAACAGATGTGTGTGTCTCGGCTCT 316
QY 311 GCGAGCGCGCGGCGCGCGCGCTTATATGGGCGAGCGCATCTTCGCGCGCGATGCGCATC 370
DB 317 CCGAGCGCGAGCGCGCGCGCGCTTTACATCGCGCATGATCATCGCGCGCGCGCGA 376
QY 371 TGAATCGCGCGCGCGCGCGCGCGCTTACGCGCGCGCGCGCGCGCGCGCGCGCGAG 430
DB 377 CCAATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAG 436
QY 431 GAGACGCGAGCCATCTCGCGGTGACAGTACCGCGCATCGGCGCGCTTGGCGCTCTGT 490
DB 437 GCGACGCGCTGCGCATCTTTCGCGGTGACAGCTTGAATGTGGCGCGCTGCGCGCTGCT 496
QY 491 GCTGGGAGCATTCAGCGCGATTCGAAATATGCGCGCGCGCGCGCGCGCGCGCGAGTCC 550
DB 497 GCTGGAGACCTGCGACCGCTCTCCAAATACCGCATGATGCGCGAAGAAAGAGTGC 556
QY 551 ACGTTCGCTGTCGCGCGAGCTTCAAGCTTATCGCGCGCATGCGCTATGCGCTCGAACCG 610
DB 557 ATGTCGCGCGCTGCGCGAGCTTTCGCTTACGATCCGTCGCGCGCGCGCTCGCGCGG 616
QY 611 AGGTCAATACCGCGCGAGCGCATATCGCGGTGCGAGCGCGCGCTGTAAGTGTGCGCT 670
DB 617 AAGTGAACATGCGCGCGAGCGAAGAAATCTATGCGGTGCGAGCGCTGTGTTTGTATCGCG 676
QY 671 CGTGGCGAGCGTTTTCGCGCGAGATGATCAAGGTATTTGATGATACCGCGCGAAGAGAG 720
DB 677 CGTGGCGAGCGTTTTCGCGCGAGATGATCAAGGTATTTGATGATACCGCGCGAAGAGATC 726
QY 731 TGTTCCTCAAGCGCGCGCGCGGCTTTCGCGCGAGTATTTTCGCGCGCGCGCGCGCTG 790
DB 737 AGTTCCTGATGCGCGCGCGCGGCTTTCGCGCGAGTATTTTCGCGCGCGCGCGCGCTG 796
QY 791 CCGAGCGCGCTCCCGAGAGCCGAAAGAGGAGTGTGTGTCGCGATATCGACTTGGCAGTGA 850
DB 797 CGGCGCGCGCTCCCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGTGA 856
QY 851 TCGCGTTGGCGAGCGCGCGCGCGATTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 910
DB 857 TTTTCGTTGCGAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 916
QY 911 GCGTGTCTGTGATGATCGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAT 970
DB 917 GCGTGTCTGTGATGATCGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGAT 976
QY 971 CGCAAAAGAGAGCAAGGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1030
```

```
DB 977 CGAAGATCCGCGCGCGCGGTGACGCCCGCGCGGTCTCTGATCTGACCGCACAG 1036
QY 1031 CCGCGCGAG 1038
DB 1037 TCCGCGCGG 1044
```

## RESULT 13

ADH36045  
ID ADH36045 standard; DNA; 1062 BP.

AC ADH36045;

DT 11-MAR-2004 (first entry)

XX Chemical process monitoring-related nitrilase gene sequence SeqID211.

XX Chemical process monitoring; biochemical process monitoring; cyanide;

KM high throughput system; gene; ds.

OS Unidentified.

PN W02003098187-A2.

XX 27-NOV-2003.

XX 15-MAY-2003; 2003MO-US015639.

XX 15-MAY-2002; 2002US-0380737P.

XX (DIVE-) DIVERSA CORP.

XX Weiner D, Chaplin JA, Chi E, Milan A, Desantis G, Burk MJ;

XX Mcquaid J, Stege J;

XX WPI; 2004-142708/14.

XX P-PADB; ADH36046.

XX Claim 74; SEQ ID NO 211; 277bp; English.

XX This invention relates to a novel method of monitoring chemical or

CC biochemical processes. The method involves providing a reactant

CC comprising cyanide (or a material that can be converted to a cyanide)

CC that generates as a reaction product cyanide or a material that can be

CC converted to cyanide and measuring the concentration of produced cyanide.

CC The method is useful for monitoring a chemical or biochemical process.

CC The method is effective for high throughput systems and is sufficiently

CC sensitive to detect a small amount of product. The present sequence is

CC that of a gene which encodes a nitrilase enzyme which can be used in the

CC method of the invention.

XX Sequence 1062 BP; 186 A; 355 C; 341 G; 180 T; 0 U; 0 Other;

XX Query Match 46.7%; Score 486; DB 12; Length 1062;

Best Local Similarity 67.8%; Pred. No. 2,3e-87;

Matches 697; Conservative 0; Mismatches 325; Indels 6; Gaps 1;

137 TTCGAGAGACTGGATACCGGGTTACCCCTGTGGATCGGCTAGGCGCGCGCTGGG 196  
QY 197 GCAT-----GCGCTTGTCCAGGGCTATTTCGAGAAATTCCTGTGTGGCGCGCAAGC 250  
DB 197 CTATCATGCGCGGCTTGTCTCGCGCTATTTCGACAACTCCCTCAGCTACCAAGACCGCGC 256  
QY 251 AGTGGCAGGCGCTGCGGATGCGGCGCGCGCGCAAGGATCATGTGTGTGGCGGCTATA 310  
DB 257 AGCGCGAGAAAGCTCCGCGCGCGCGCGCAAGCGCAAGATGATGTGTGTGTGTGTGTGT 316  
QY 311 GCGAGCGCGCGCGCGCGCGCGCTTATATATGAGCGCGCGCATCTTGTGGCGCGCGCATC 370  
DB 317 CCGAGCGCGCGAGCGCGCGCGCTTTACATCGCGCATGATCATCGCGCGCGCAAA 376  
QY 371 TGATCGCGCGCGCGCGCAAGCTCAAGCTTACCTCAATCGCGAGCGCAAGCTGTTCGCGAG 430  
DB 377 CCATCGCGCAAGCGCGCGCAAGCTCAAGCGCGCAAGCGCGAGCGGAGCGGTGTTCGCGAG 436  
QY 431 GAGAGCGAGCGCATCTGCGCGGTGACCATACCGCATGCGCGCGCTGCGCGCGCTGT 490  
DB 437 GCGAGCGCTCGCATCTTGTGGTGAACGAGCTTGATGTGTGTGTGTGTGTGTGTGTGT 496  
QY 491 GCTGGAGCAGCATCAGCGCATTTGTGAAATAGCGCATGTACGCGCGCGCAAGAGGTCC 550  
DB 497 GCTGGAGCAGCATCAGCGCGCTGTCTCAAAATAGCGCATGTATGCGAGAGACAGAGTGC 556  
QY 551 ACGTGGCTGTGTGCGCGAGCTTCAAGCTTATCGCGCGCATGCGCTTGTGAGCGG 610  
DB 557 ATGTGCGCGCGCTGCGCGAGCTTTTGTGCTTTCAGATCCGCTTGCAGCAAGCGCTGCGCGG 616  
QY 611 AGGTCAATACCGCGCGCAAGCGCATCTACGCGGTGAGAGGCGCGCTGCTACGCTGCGGT 670  
DB 617 AAGTGAATATGCGCGAGCAAAATCTATGCGGTGAGAGGCTGCTGTGTGTGTGTGTGT 676  
QY 671 CGTGGCGAGCGCTTTCGCGGAGATGATCAAGTATTGTGATACGCGCGCAAGAGAGA 730  
DB 677 CGTGGCGAGCGCTTTCGCGGAGATGATCAAGTATTGTGATACGCGCGCAAGAGAGATC 736  
QY 731 TGTTCCTCAAGCGCGCGCGGCTTTTGTGATGATTTTGTGCGCGCGCGCGCGCGCTGG 790  
DB 737 AGTTCCTGATGCGCGCGCGCGGCTTTGTGATGATTTTGTGCGCGCGCGCGCGCGCTCG 796  
QY 791 CCGAGCGCGCTCCCGAGACCGAAGAGGATGCTGTGTGTGTGTGTGTGTGTGTGTGTGT 850  
DB 797 CCGGCGCGCTGCGCGCGCGCGCAAGAGGCTTGTCTTACGCGCATTCGATTCGCGGATGA 856  
QY 851 TCGGCTTGGCAAGCGCGCGCGCGCGATTCGCGCGCGCACTTATTCAGCGCGCGAGTAAAGC 910  
DB 857 TTTTGTGTTGCCAAGCGCGCGCGCGATTCGCGCGCGCGCATTTATGACGCGCGCGCATCC 916  
QY 911 GGTGCTGTGATCGATCGATCGCGCGCAAGCGGTGTGTGTGTGTGTGTGTGTGTGTGT 970  
DB 917 GGTCTTGTGTCAACATCGCGCTGTGGTATTCGGGTGTGAGACCATGCGGTGTGCGATGATG 976  
QY 971 CCGAAAGAGAGACAAAGGCGAGCGCGCGCGCGCGCGGTGTGTGTGTGTGTGTGTGTGT 1030  
DB 977 CCGAGACCAAGCGAGAGCAACCGGTATGACCGGAACCAAGGCAACCGAAGCGTGTGCGCGT 1036  
QY 1031 CCGGCGAG 1038  
DB 1037 TCGCGCGG 1044

RESULT 14  
ADG93746  
ID ADG93746 standard; DNA; 1062 BP.

XX ADG93746;  
XX AC  
XX 11-MAR-2004 (first entry)  
XX DE Nitrilase enzyme gene sequence Segid211.

KW nitrilase; nitrile; carboxylic acid; chemical process; pH; temperature;  
KW enantioselective transformation; gene; ds.  
OS Unidentified.  
PN MO2003097810-A2.  
XX 27-NOV-2003.  
XX 15-MAY-2003; 2003MO-US015712.  
XX 15-MAY-2002; 2002US-00146772.  
PR 09-SEP-2002; 2002US-00241742.  
XX (DIVE-) DIVERSA CORP.  
PA Desantis G, Short JM, Burk M, Wong K, Farwell R, Chatman K;  
PI WPI; 2004-090637/09.  
DR P-PSDB; ADG93747.  
XX  
XX New isolated or recombinant nucleic acid encoding a polypeptide having  
PT nitrilase activity, useful for screening enantioselective transformation.  
PS Claim 44; SEQ ID NO 211; 295pp; English.  
XX  
XX This invention is related to a novel isolated or recombinant nucleic acid  
CC encoding a protein having nitrilase activity. Nitrilase's are capable of  
CC converting nitrile's directly to carboxylic acids and have great  
CC potential for use in industrial chemical processes. The isolated  
CC nitrilase proteins of the invention have increased activity and stability  
CC at increased pH and temperature when compared to those conventionally  
CC used. In addition, the nucleic acid of the invention is useful for  
CC screening enantioselective transformation. The present sequence is that  
CC of a DNA sequence which encodes a nitrilase enzyme of the invention.  
XX  
SQ Sequence 1062 BP; 186 A; 355 C; 341 G; 180 T; 0 U; 0 Other;  
Query Match 46.7%; Score 486; DB 12; Length 1062;  
Best Local Similarity 67.8%; Pred. No. 2.3e-87;  
Matches 697; Conservative 0; Mismatches 325; Indels 6; Gaps 1;  
QY 17 CGAAGTATCGCGCGCGCGGTGAGCGCGCGCGCGGTCTTGTGATCGACCGCAG 76  
DB 17 CGAAGTATCGAAGTATCGCGCGCGGTGAGCGCGCGCGCGGTCTTGTGATCGACCGCAG 76  
QY 77 TCGAAGAGCGATCGCGCTGATCGAGCGCGCGCAAGAGAGTGTGTGTGTGTGTGTGTGT 136  
DB 77 TCGAAGAGCGCTGCTTGTATCGACGAGCGCGCGCGCGCGCGCGCGCGCTTATCGCCT 136  
QY 137 TCCGAGACTTGTGATTCGCGCTATCCCTTTTGTGATATGAGTGTGTGTGTGTGTGTGT 196  
DB 137 TTTCCGAGACTTGTGATTCGCGCTATCCCTTTTGTGATATGAGTGTGTGTGTGTGTGTGT 196  
QY 197 GCAT-----GCGTGTGTGCAAGCGCTATTTCGAGAAATTCCTGTGTGTGTGTGTGTGT 250  
DB 197 CTATCATGCGCGGCTTGT 256  
QY 251 AGTGGCAGGCGCTGCGGATGCGGCGCGCGCGCGCAAGGATCATGTGTGTGTGTGTGTGT 310  
DB 257 AGCGCGAGAAAGCTCCGCGCGCGCGCGCGCAAGCGCAAGATGATGTGTGTGTGTGTGT 316  
QY 311 GCGAGCGCGCGCGCGCGCGCGCTTATATATGAGCGCGCGCATCTTGTGGCGCGCGCATC 370  
DB 317 CCGAGCGCGAGCGCGCGCGCGCTTTATACATCGCGCATGATCATCGCGCGCGCAAA 376  
QY 371 TGATCGCGCGCGCGCGCAAGCTCAAGCTTACCTCAATCGCGAGCGCAAGCTGTTCGCGAG 430  
DB 377 CCATCGCGCAAGCGCGCGCAAGCTCAAGCGCGCAAGCGCGAGCGGAGCGGTGTTCGCGAG 436  
QY 431 GAGAGCGAGCGCATCTGCGCGGTGACCATACCGCATGCGCGCGCTGCGCGCGCTGT 490  
DB 437 GCGAGCGCTCGCATCTTGTGGTGAACGAGCTTGATGTGTGTGTGTGTGTGTGTGTGT 496



QY 491 GCTGGAGACATCCAGCATTTGCAAAATACGCCATGTACGCCCGCAGCAAGTCC 550  
Db 497 GCTGGAGAACCTCCAGCCCGCTGCCAAATACGCCCATGTATGCCAGAACCAAGGTGC 556  
QY 551 ACCTGCGGTGTCGCGAGCTTTCAGCTCTATCCGGGATGCGCTATGCGCTCGGACCGG 610  
Db 557 ATGTGCGGCGCTGGCCAGCTTTTCGCTTTACGATCCGTTGCGCAGCGCTCGGCGCGG 616  
QY 611 AGGTCAATACCGCCGCAAGCCAGATCTACCGGCTGCAAGGCGGCGTGTACGTCTGCGT 670  
Db 617 AAGGAACAATGCGCGCAGCAAAATCTATGCGGTGCAAGGCTCGTGTTCGTTCATCCGCG 676  
QY 671 CGTCCGCGACCGTTTCGCGCGAGATGATCAAGGATTGATGATGATGATGATGATGATGAT 730  
Db 677 CGTCCGCGACCGTTTCGCGCGAGATGATCAAGGATTGATGATGATGATGATGATGATGAT 736  
QY 731 TGTTCCTCAAGAGCGCGCGCGGCTTTTGCATGATTTTCGCGCGCGCGCGCGCGCTTGG 790  
Db 737 AGTTCTGCAATGCGCGCGCGCGGCTTTTGCATGATTTTACGCGCGCGCGCGCGCGCTTGG 796  
QY 791 CCGAGCGGCTTCCCGGAGACCGGAAGAGGACTGCTGTGCGCGGATATGACTCTCGGCTGA 850  
Db 797 CCGGCGCGCTGCG 856  
QY 851 TCGCGTTGCGCAAGCG 910  
Db 857 TTTTCGTTGCGCAAGCG 916  
QY 911 GCGTCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 970  
Db 917 GCGTCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 976  
QY 971 CGCAAAACGAGGACCAAGGCGGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1030  
Db 977 CGAGAGCAAGCGCGGAGGACCGCGCTTACCGCGAAGCAAGGACCGAAGCGTGGCGCGT 1036  
QY 1031 CCGCGCAG 1038  
Db 1037 TCGCGCGG 1044

RESULT 15  
AD162343  
ID AD162343 standard; DNA; 1062 BP.  
AC AD162343;  
XX  
DT 22-APR-2004 (first entry)  
DE DNA encoding nitrilase polypeptide #106.  
XX  
XX Acorvastatin; (R)-ethyl 4-cyano-3-hydroxybutyrate;  
XX (R)-ethyl 4-cyano-3-hydroxybutyric acid; epichlorohydrin;  
KM 3-hydroxyglutaronitrile; 3-hydroxyglutaronitrile;  
KM 4-cyano-3-hydroxybutyric acid; ethyl-4-cyano-3-hydroxybutyric acid;  
KM mixed hyperlipidaemia; homozygous familial hypercholesterolaemia;  
KW antilipemic; gene; ds.  
XX  
XX Unidentified.  
OS  
XX  
XX W02003106415-A2.  
PN  
XX  
XX 24-DEC-2003.  
PD  
XX  
XX 13-JUN-2003; 2003WO-US018840.  
PF  
XX 13-JUN-2002; 2002US-0389317P.  
PR 28-JUN-2002; 2002US-0392944P.  
XX  
XX (DIVE-) DIVERSA CORP.  
PA  
XX  
XX Burk M, Desantis G, Morgan B, Zhu Z;  
PI

XX WPI; 2004-090821/09.  
DR P-FSDB; AD162344.  
DR  
XX  
PT Preparation of atorvastatin comprises catalytic conversion of 3-  
PT hydroxyglutaronitrile by polypeptide with nitrilase activity, converting  
PT obtained 4-cyano-3-hydroxybutyric acid to ethyl-4-cyano-3-hydroxybutyric  
PT acid and forming atorvastatin.  
XX  
XX Claim 46; SEQ ID NO 211; 253bp; English.  
PS  
XX  
XX The present invention relates to a method for preparing an atorvastatin  
CC intermediate known as (R)-ethyl 4-cyano-3-hydroxybutyrate ((R)-ethyl 4-  
CC cyano-3-hydroxybutyric acid). The method comprises optionally converting  
CC epichlorohydrin or equivalent to 3-hydroxyglutaronitrile, catalytic  
CC conversion of 3-hydroxyglutaronitrile or equivalent to 4-cyano-3-  
CC hydroxybutyric acid with a polypeptide having nitrilase activity,  
CC converting 4-cyano-3-hydroxybutyric acid to ethyl-4-cyano-3-hydroxybutyric  
CC acid, and converting this to (R)-ethyl 4-cyano-3-hydroxybutyrate. The  
CC method involves whole cell processes, cell lysate process, "one pot"  
CC processes, and "multi-pot" processes using a variety of parameters.  
CC Acorvastatin is used, in conjunction with dietary restriction, in the  
CC management of hyperlipidaemia, including hypercholesterolaemia, mixed  
CC dyslipidaemia and homozygous familial hypercholesterolaemia. The present  
CC sequence encodes a nitrilase polypeptide obtained from an environmental  
CC sample.  
XX  
SQ Sequence 1062 BP; 186 A; 355 C; 341 G; 180 T; 0 U; 0 Other;

Query Match 46.7%; Score 486; DB 12; Length 1062;  
Best Local Similarity 67.8%; Pred. No. 2.3e-87;  
Matches 697; Conservative 0; Mismatches 325; Indels 6; Gaps 1;

QY 17 CGAAGTATCG 76  
Db 17 CGAAGTATCGAAGTATCG 76  
QY 77 TCGAGAAAGCGATCGGCTGATCGAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 136  
Db 77 TCGAGAAAGCGATCGGCTGATCGAGCAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 136  
QY 137 TCCGAGAGCTTGGATTCGCGGCTATCCCTTTTGGATATGCTGGCGCGCGCGCGCTTGG 196  
Db 137 TCCGAGAGCTTGGATTCGCGGCTATCCCTTTTGGATATGCTGGCGCGCGCGCGCTTGG 196  
QY 197 GCAT-----GCGCTTGTTCAGCGCTATTCGAGATTCGCTGCGCGCGCGCGCGCG 250  
Db 197 GCAT-----GCGCTTGTTCAGCGCTATTCGAGATTCGCTGCGCGCGCGCGCGCGCG 250  
QY 251 AGTGGAGGCGCTGCGGATGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTATA 310  
Db 257 AGGCGAGAAAGCTCCGCTCT 316  
QY 311 GCGAGCG 370  
Db 317 CCGAGGCG 376  
QY 371 TGATCCCG 430  
Db 377 CCATCCCG 436  
QY 431 GAGAGCGAGCGATCTCGCGGTGCGAGATCGCGCATCGGCGCGCTCGCGCGCGCTCTGT 490  
Db 437 GCGAGCGCTCGGATCTTGGCGGTGCGAGATCGGCTGATGTTGGCGCGCTCGCGCGCTGT 496  
QY 491 GCTGGAGACATCCAGCATTTGCAAAATACGCCATGTACGCCCGCAGCAAGTCC 550  
Db 497 GCTGGAGAACCTCCAGCCCGCTGCCAAATACGCCCATGTATGCCAGAACCAAGGTGC 556  
QY 551 ACCTGCGGTGTCGCGAGCTTTCAGCTCTATCCGGGATGCGCTATGCGCTCGGACCGG 610  
Db 557 ATGTGCGGCGCTGGCCAGCTTTTCGCTTTACGATCCGTTGCGCAGCGCTCGGCGCGG 616





**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 26, 2006, 22:54:39 ; Search time 214.785 Seconds  
(without alignments)  
8615.305 Million cell updates/sec

Title: US-09-751-299-1

Perfect score: 1041

Sequence: 1 atgcgcgagccatgcagaa.....gcgcgcgcgcgcgcagtag 1041

Scoring table: IDENTITY\_NUC

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0\*

Maximum Match 100\*

Database :

- Issued Patents NA:\*
- 1: /cgn2\_6/ptodata/1/ina/1 COMB.seq:\*
  - 2: /cgn2\_6/ptodata/1/ina/5 COMB.seq:\*
  - 3: /cgn2\_6/ptodata/1/ina/6A COMB.seq:\*
  - 4: /cgn2\_6/ptodata/1/ina/6B COMB.seq:\*
  - 5: /cgn2\_6/ptodata/1/ina/H COMB.seq:\*
  - 6: /cgn2\_6/ptodata/1/ina/PC/US COMB.seq:\*
  - 7: /cgn2\_6/ptodata/1/ina/PP COMB.seq:\*
  - 8: /cgn2\_6/ptodata/1/ina/RE COMB.seq:\*
  - 9: /cgn2\_6/ptodata/1/ina/Backfile1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	343.2	33.0	1035	2	US-08-690-493-2	Sequence 2, Appli
2	343.2	33.0	1200	2	US-08-690-493-3	Sequence 3, Appli
3	297.4	28.6	1071	3	US-09-806-876A-1	Sequence 1, Appli
4	283.8	27.3	1110	3	US-09-823-373-4	Sequence 4, Appli
5	283.8	27.3	1110	3	US-09-823-373-13	Sequence 13, Appli
6	283.8	27.3	1776	3	US-09-823-373-15	Sequence 15, Appli
7	202	19.4	1110	3	US-09-823-373-16	Sequence 16, Appli
8	132.2	12.7	385	3	US-09-823-373-3	Sequence 3, Appli
9	109.6	10.5	1194	2	US-08-447-702-4	Sequence 4, Appli
10	108	10.4	1194	2	US-08-447-702-4	Sequence 4, Appli
11	57	5.5	1995	3	US-09-902-540-4902	Sequence 4902, Ap
12	57	5.5	28493	3	US-09-902-540-1241	Sequence 1241, Ap
13	55	5.3	1728	3	US-09-758-759-126	Sequence 126, App
14	55	5.3	109519	3	US-09-758-759-1	Sequence 1, Appli
15	54.6	5.2	5877	3	US-10-152-886-94	Sequence 94, Appli
16	51.8	5.0	58857	3	US-09-477-962-1	Sequence 1, Appli
17	51.6	5.0	507	3	US-09-902-540-4718	Sequence 4718, Ap
18	51.6	5.0	5121	3	US-09-902-540-704	Sequence 704, App
19	51.6	5.0	28762	3	US-09-902-540-1232	Sequence 1232, App
20	51.4	4.9	2124	3	US-09-902-540-7586	Sequence 7586, App
21	51.4	4.9	7704	3	US-09-902-540-743	Sequence 743, App
22	50.8	4.9	105413	3	US-10-427-923-3	Sequence 3, Appli
23	50.8	4.9	112219	3	US-09-949-016-12453	Sequence 12453, A
24	50.8	4.9	112222	3	US-09-949-016-14324	Sequence 14324, A

C	25	50.4	4.8	109519	3	US-09-758-759-1	Sequence 1, Appli
	26	49.8	4.8	30001	2	US-08-125-468-1	Sequence 1, Appli
	27	49.8	4.8	30001	2	US-08-474-933-1	Sequence 1, Appli
	28	49.4	4.7	1377	3	US-09-252-991A-10413	Sequence 10413, A
	29	49.4	4.7	1377	3	US-09-252-991A-10306	Sequence 10306, A
	30	49.4	4.7	1410	3	US-09-252-991A-10749	Sequence 10749, A
	31	48.8	4.7	1494	3	US-09-902-540-2363	Sequence 2363, Ap
	32	48.8	4.7	15447	3	US-09-902-540-1100	Sequence 1100, Ap
	33	48.6	4.7	1088	3	US-09-252-991A-14407	Sequence 14407, A
	34	48.6	4.7	1566	3	US-09-252-991A-14264	Sequence 14264, A
	35	48.6	4.7	1508	3	US-09-252-991A-14353	Sequence 14353, A
	36	48.4	4.6	699	3	US-09-252-991A-4020	Sequence 4020, Ap
	37	48.2	4.6	1377	3	US-09-377-557-17	Sequence 17, Appli
	38	47.6	4.6	939	3	US-09-902-540-6060	Sequence 6060, App
	39	47.6	4.6	941	3	US-09-902-540-252	Sequence 252, App
	40	47.6	4.6	2838	3	US-09-252-991A-14128	Sequence 14128, A
	41	47.6	4.6	3897	3	US-09-252-991A-14008	Sequence 14008, A
	42	47.4	4.6	3957	3	US-10-237-551-193	Sequence 193, App
	43	47.4	4.6	23673	3	US-09-773-816-1	Sequence 1, Appli
	44	47.4	4.6	154746	3	US-09-827-688-8	Sequence 8, Appli
	45	47.4	4.6	154746	3	US-09-827-688-8	Sequence 8, Appli

ALIGNMENTS

RESULT 1  
US-08-690-493-2  
Sequence 2, Application US/08690493  
Patent No. 5872000  
GENERAL INFORMATION:  
APPLICANT: Yu, Fujio  
TITLE OF INVENTION: No. 5872000el Nitrlase Gene  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Steinberg, Raekin & Davidson, P.C.  
STREET: 1140 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: MS-DOS Editor  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/690,493  
FILING DATE: 31 JUL-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 213061/1995  
FILING DATE: 31-JUL-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Davidson, Clifford M  
REGISTRATION NUMBER: 32,728  
REFERENCE/DOCKET NUMBER: 3821005  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 768-3800  
TELEFAX: (212) 382-2124  
INFORMATION FOR SEQ ID NO: 2:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1035 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: double  
TOPOLOGY: linear  
MOLECULE TYPE: genomic DNA  
ORIGINAL SOURCE:  
ORGANISM: Gordana terrae  
STRAIN: MA-1  
CELL TYPE: unicellular organism  
US-08-690-493-2  
Query Match 33.0%; Score 343.2; DB 2; Length 1035;

Best Local Similarity 61.8%; Pred. No. 1.5e-67;  
Matches 546; Conservative 0; Mismatches 338; Indels 0; Gaps 0;

```
QY 31 GCGCGCGTGCAGCGCGCGCGTCTCTCGATCTCGACCGCAGTCGAGAAAGCATC 90
DB 34 GCGGTGACCCGCGCGCAACCGGTGTGCTTCGACCTCTCGCGCACCGTCGACAGACCATC 93
QY 91 GCGCTGATCGAGCGCGCGCGCAACGATGCGCGCTGATCGCATTCGAGACATTTGG 150
DB 94 GCGCTGCTGAGAGCGCTCCCGGCGCGCGCGCATCGATCGCGGTTCCGGAACCTCG 153
QY 151 ATTCGCGCTATCCCTTTTGGATATGCTGCGCGCGCGCGCTTGGCGATCGCTTCTC 210
DB 154 ATACCGGGGTACCGGTGCTTCTGTGCTTGAATCGGTGCGCTGCGAGACGACATCTTC 213
QY 211 CAGCGCTATTTGAGAAATTCGCTGTGCGCGCGCGAGCAAGCATGTGCGAGCCCTTGG 270
DB 214 ATCCGATATCGCGAAGCTCGCTCGATCTCGACGCGAGCTTCGCGCGCATCAGGAA 273
QY 271 GCGCGCGCGCGCGCGCATGCTGCTGCGCGCTATAGCAGCGCGCGCGCGAGC 330
DB 274 GCGCGCAAGAAAGACGACATGCGCATCACCATGGGATTTAGTGAGCCCGTCACTTCG 333
QY 331 CTCTATATGGGCGCGCGATTTTGGCGCGCGATGCGCATCTGATCGCGCGCGCGAAG 390
DB 334 CTGTACATGGGCGCGCGGTCTGATGAGCGGTGCTGCTACGCGACACGCGCGCAA 393
QY 391 CTCAGGCTTACCGATGCGGAGCGCGCGTGTGCGCGAGGAGCGCGAGCATCTGCG 450
DB 394 CTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 453
QY 451 GTGCGAGATACCGCGCATCGCGCGCGCGCGCGCTGTGCTGCGAGCAATCCAGCA 510
DB 454 GTGAGCCGAGCAAGTCTCGCGCGAGTGGGTGCTGTGCTGTGGAAATCTTCAACCG 513
QY 511 TTGTGCAATACCGCATGCTACGCGCGCGCGCAAGAGTCCACGCTGCGCTGTGCGCG 570
DB 514 TTGACCAAGTACCGCATGCTACGCGCGCGCGCAAGAGTCCACGCTGCGCTGTGCGCG 573
QY 571 TTGACCTATCGCGCGCATGCTGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCG 630
DB 574 TTCTCGATCTTCCCGCGCGCGGTGTATGCGTCCGCGCGCGCGCGCGCGCGCTCT 633
QY 631 CAGATCTACCGCGTCTGAGCGCGCGCTGATGCTGCGCGCGCGCGCGCGCGCGCG 690
DB 634 CAGCAATACCGCGTACGAGCGCGCGCAACCTGCTCTCGCATGCGCGGTCTATCGCG 693
QY 691 GAGATGATCAAGGATATGCTGATAGCGCGCGCAAGAGATGTTCTCAAGCGCGCGCG 750
DB 694 GCGGTGCGAGCGGTTCGCGATACCGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 753
QY 751 GGTTCGCGATGATTTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 810
DB 754 GGTATGCGCGTATCTACGCGTCCGAGCGGTCTGCTCACTCGCGAAGCGGTCTCGCG 813
QY 811 GAAGAGGAGAGCTGCTGCTGCGCGATGAGCATCTCGCGCATGATCGGTGCGCGAG 870
DB 814 GAGAGGAGAGATCTGTAAGCGCGAGATGATCTGTCTGCGATCTGCGCGCAAGAG 873
QY 871 GCGGATCGCGCGCGCGCATATTCAGCGCGCGCGAGTAAAGCGCGCT 914
DB 874 GCGGACCGCGTGGGACATCTACGCGGTCTGCGAGGATGCTGCGTCT 917
```

RESULT 2  
US-08-690-493-3

; Sequence 3, Application US/08690493

; Patent No. 5872000

; GENERAL INFORMATION:

; APPLICANT: Yu Fujio

; TITLE OF INVENTION: No. 5872000e1 Nitrlase Gene

; NUMBER OF SEQUENCES: 5

; CORRESPONDENCE ADDRESS:

ADDRESSEE: Steinberg, Raekin & Davidson, P.C.  
STREET: 1140 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: MS-DOS Editor

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/690,493

FILING DATE: 31 JUL-1996

PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP 213061/1995

FILING DATE: 31 JUL-1995

ATTORNEY/AGENT INFORMATION:

NAME: Davidson, Clifford M

REGISTRATION NUMBER: 32,728

REFERENCE/DOCKET NUMBER: 3821005

TELEPHONE: (212) 768-3800

TELEFAX: (212) 382-2124

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 1200 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: genomic DNA

ORIGINAL SOURCE:

ORGANISM: Gordona terrae

STRAIN: MA-1

CELL TYPE: unicellular organism

US-08-690-493-3

Query Match 33.0%; Score 343.2; DB 2; Length 1200;  
Best Local Similarity 61.8%; Pred. No. 1.6e-67;  
Matches 546; Conservative 0; Mismatches 338; Indels 0; Gaps 0;

```
QY 31 GCGCGCGTGCAGCGCGCGCGTCTCTCGATCTCGACCGCAGTCGAGAAAGCATC 90
DB 128 GCGGTACCGCGCGCGCGCGCGGTGTGCTTCGACCTCTCGCGCACCGTCGACAAAGCAT 187
QY 91 GCGCTGATCGAGCGCGCGCGCAAGCAGAGCGTGGCGCTGATCGCATTTCCAGAGACTTG 150
DB 188 GCGCTGATCGAGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 247
QY 151 ATTCGCGCTATCCCTTTTGGATATGCTGCGCGCGCGCGCGCGCGCGCGCGCG 210
DB 248 ATACCGGGGTACCGCGGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 307
QY 211 CAGCGCTATTTGAGAAATTCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 270
DB 308 ATCCGATATCGCGAGATCGCTGCTGATCTGACGCGCGCGAGTTCGCGCGCATCAGGAA 367
QY 271 GCGCGCGCGCGCGCGCGCATGATGCTGCGCGCGCGCTATAGCAGCGCGCGCGCAGC 330
DB 368 GCGCGCAAGAAAGACATCGCGATCACCATGGATTTAGTAGAGCGGTGATGTTTG 427
QY 331 CTCTATATGGGCGCGCGCATTTTCGCGCGCGCGCGCGCGCGCGCGCGCGCG 390
DB 428 CTGTACATGGGCGCGCGCGCATTCAGCGTGAAGCGGTGCTGTACGACACGCGCGCAA 487
QY 391 CTCAGCTTACCGCATGCGCGCGCGCGCGGTTCGCGCGAGAGAGAGCGCGCGCATCTCG 450
DB 488 CTGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 547
QY 451 GTGACGATACCGCGCATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 510
DB 548 GTGACGACGACGATCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 607
```

```
QY 511 TTGTGAAATACGCGATGATACGCGCCGAGACAGAGTCCAGTCCGCGTGGCCGAGC 570
DB 608 TTGACCAAGTACGCGCATCTACTCCGACAGACAGATTCCATCCGCGCATGGCCAGC 667
QY 571 TTGAGCTCTATCGCGGATGCGCTATGCTCGAGCCGAGGTCATATCCGCGCAAGC 630
DB 668 TTTCGATCTTCCGCGGCGCGGTGTATGCGCTCGGGCCGAGGTCAACACGCGGCTCT 727
QY 631 CAGATCTACGCGGTGAGAGGCGGCTGTAAGTGTGCGGTGCGCGGACGCTTTCCGCG 690
DB 728 CAGCAATAACGCGCTGAGAGGCGAGACCTACCTTCGCTCATGCGCGGTATCGCGAT 787
QY 691 GAGATGATCAAGGATTTGATGATATACGCGCCGACAAAGAGATGTTCTCAAGCCGCGCGC 750
DB 788 GCAGGTGGAGAGGCGTTTGCCGATACGAGAGAAAGCGACCTCATCACAAGAGAGGC 847
QY 751 GGTTTGCGATGATTTTTCGAGCCCGACGCGCCGCTGCGCGAGCGCTCCGAGAC 810
DB 848 GGATATGCGCGTATCTAGCGGTCCGACGCGTTCCTGCGGAAACGCGTCCGCGCAT 907
QY 811 GAAGAGGAGTGTGCTGCTGCGCGATATGCACTCGCGATGATCGGTTGGCCAGCGCG 870
DB 908 GACGAGGGAATCTGTACGCGGACATCGATCTGTCTGCGATTTGCGCGCAAGAACCCG 967
QY 871 GCGGATCCGCGGCGCACTATTTCAGCGCGCGACGTAACGCGGCT 914
DB 968 GCGGACCGGCTTGAGCACTACTCGGTCCGAGCGTACTGCGCT 1011
```

## RESULT 3

```
US-09-806-876A-1
; Sequence 1, Application US/09806876A
; Patent No. 6869783
; GENERAL INFORMATION:
; APPLICANT: Rees-Loeschke, Marion
; APPLICANT: Friedrich, Thomas
; APPLICANT: Hauser, Bernhard
; TITLE OF INVENTION: A process for preparing chiral carboxylic acids
; TITLE OF INVENTION: from nitriles using a nitrilase or microorganisms
; FILE REFERENCE: 49462
; CURRENT APPLICATION NUMBER: US/09/806,876A
; PRIOR FILING DATE: 2001-05-21
; PRIOR APPLICATION NUMBER: Germany/19848129.2
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: WordPerfect version 6.1
; SEQ ID NO 1
; LENGTH: 1071
; TYPE: DNA
; ORGANISM: Alcaligenes faecalis
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1... 1071
US-09-806-876A-1
```

```
Query Match 28.6%; Score 297.4; DB 3; Length 1071;
Best Local Similarity 58.2%; Pred. No. 2,6e-57;
Matches 523; Conservative 0; Mismatches 376; Indels 0; Gaps 0;
```

```
QY 25 CGCGCGCGCGGTCGAGCGCGCGCGCGGTTCCTTCATCTCGACCGCACAGTGGAGAAA 84
DB 22 CGGCGAGCGCGCTGACAGCGCGCGCTCTCCAACTACGATCTGGCAACGCGGTGGATAAA 81
QY 85 GCGATCGGCTGATCGAGCGCGCGCAAGACAGCGCTGATCGCATTTCCAGAG 144
DB 82 ACATTTAGCTGCTGCTGCAAGCGCGCGATGAGGGCTGTAGCTGATGTGTTGGTGA 141
QY 145 ACTTGATTTCCGCGCTATCCCTTTTGTGATATGCTGGCGCGCGCTTGGGAGATGCGC 204
DB 142 ACTTGCGTCCCGGATATCCCTTCACAGCTGCGGTGGCGCACGCGCGCTGGTCCGTA 201
QY 205 TTGCTCCAGCGCTATTTTCGAATTTCTGCTGCGCGCGCACAGCAAGTGGACGCGCTG 264
```

```
DB 202 TACAGTCCCGCTACTATGCAACTGCGCTCTCGCTGAGACAGTGAAGTTTCAACGATT 261
QY 265 GCGGATCGGCGCGCGCGCAACGCGATGATGTGTGCGCGCTATGACGAGCGCGCGC 324
DB 262 GCCAGGCGCGCAAGACTTGGGTATTTTTCATCGCACTGGGTATATGCGAGCGCGC 321
QY 325 GCGAGCTCTATATGCGGCGAGGAGTCTTGGGCGCGATGCGGATCTGATCGCGCGCGC 384
DB 322 GCGAGCTTTTACCTGGGCGAATGCTGATGAGCAAGAGGCGAGATGCTGTGTCGCT 381
QY 385 CGCAAGCTCAAGCTTACCCATGCGAGCGCACCGCTGTTCCGCGAGGAGAGCGCAGCAT 444
DB 382 CGCAATCTCAACCCAGCGATGTAGAGCGCACCGTATTTGTGTAAGTTATGCCCGAT 441
QY 445 CTCGCGTGCAGATATCCGCTATCGCGCGCTTGGCGCGCTCTGTTGCTGGAGACATC 504
DB 442 CTGATTTGTCCGACACAGAACTGGAGCGGTGCTGATGCTGCTGGAGCATTTG 501
QY 505 CAGCGATTTGCAAAATCGCGCATGTACGCGCGCGACGCAAGGTCCACGTGCGTGG 564
DB 502 TCGCCTTGAGCAAGTACGCGCTGTACTCCAGCATGAACCATTCACATTTGCTGG 561
QY 565 CCGAGCTTCAGCTCTATCGCGCATGCGCTATGCGCTCGAGCCGAGGTCATTAACGCC 624
DB 562 CCGTCTTTTCTGCTATACAGCGAAGAGCCGCGCTCAGTGCAGAGTGAACATGGCT 621
QY 625 GCAAGCCAGATCTACCGCGTGCAGAGCGCGCTGCTACGTGCGTGCAGCGCAT 684
DB 622 GCTCGCAATCTATTCGTTGTAAGGCGAGTGTTCATTCGCGCGCACAGTGTGTC 681
QY 685 TCGCGGAGTATGATCAAGGATTTGATGAGTACCGCGCGACGCAAGAGATTTCTCAAGCC 744
DB 682 ACCCAAGAGCGTACGACATGCGAGAGTGGTGAACACACACCGCCCTTGTAAGTG 741
QY 745 GCGCGCGTTTTCATGATTTTTCGCGCGCGACGCGCGCGCTGCGCGAGCGCTCCG 804
DB 742 GCGCGCGCGAGTTCATGATTTTTCGCGCGCGAGCGACGACATGCTCCCTACTGCT 801
QY 805 GAGACCGAAGAGGAGCTGCTGTCGCGATATGCACTCGGCATGATCGGTTGGCCAG 864
DB 802 CAGATGCCAGAGGCTTGAATTCGCGATCTGAATATGAGAGGATTTGCTGCCAAA 861
QY 865 GCGCGCGCATTCGCGCGCGCGCTATTCACGCGCGCGCATACGCGGCTGCTGTGA 923
DB 862 GCGATCAATGACCCGTAAGGCACTATTCCAAACCGAGGCGCACCGCTGTGTCGTA 920
```

## RESULT 4

```
US-09-823-373-4
; Sequence 4, Application US/09823373
; Patent No. 6870038
; GENERAL INFORMATION:
; APPLICANT: Chauhan, Sarita
; APPLICANT: DiCosimo, Robert
; APPLICANT: Payne, Mark
; APPLICANT: Gavigan, John
; APPLICANT: Fallon, Robert
; TITLE OF INVENTION: Isolation and Expression of a Gene for Nitrilase from
; TITLE OF INVENTION: Acidovorax Facilis 72W
; FILE REFERENCE: BC-1032 US NA
; CURRENT APPLICATION NUMBER: US/09/823,373
; PRIOR FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 60/193,707
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 4
; LENGTH: 1110
; TYPE: DNA
; ORGANISM: Acidovorax facilis
US-09-823-373-4
```

Query Match 27.3%; Score 283.8; DB 3; Length 1110;  
Best Local Similarity 57.0%; Pred. No. 2.8e-54;  
Matches 519; Conservative 0; Mismatches 392; Indels 0; Gaps 0;

QY 19 AAGTATCGCGCGCGCGGTCAGAGCCGCGCGGTGTTCTCGATCTTCAGCCGACAGTC 78  
DB 19 AAGTTCCTCGCGGCAACCGTTCAAGGCAAGCGCGGTATGCTCGACGCAAGCAACATC 78  
QY 79 GAGAAAGCATCGCGCTGATGAGAGCGCGCCAGAGCAAGCGTGCCTGATCGCATTC 138  
DB 79 GACAAAGTCGATCGGATCATCGAAGAAAGCTCCCAAAAGGCGCGAGTCTGATCGCTTC 138  
QY 139 CCAGAGACTTGGATTCCCGGCTATCCCTTTGATATGAGTGGCGGCGCGCTTGGGGC 198  
DB 139 CCGAAGATTCATTTCCGGGCTACCCCTATTGGCGGTGCTCGCGACGTAAAGTACAGC 198  
QY 199 ATGCGCTTCGTCAGCGCTATTTGAGAAATTCGCTGTCGCGCGAGCAAGCATGCGAG 258  
DB 199 CTAGCTTTACTTACGCTATCAGAGAAATTCGTTGAGTGAAGTGAACCGTATCGGT 258  
QY 259 GCCCTGCGGATGGCGCCCGCGCCAGCGCATGATGCTGGCGCGCTATAGCGAGCGC 318  
DB 259 CGCTTCAGCTGGCGCGCGCGCGCAACAAATCGCACTGCTATGAGGCTATTCGAGCGG 318  
QY 319 GCGGCGCGCAGCGCTCTATATGAGCGCAGCGCATCTTCGCGCCCGATGGCGATCTGATCGCC 378  
DB 319 GAAGCGGATCGCGCTATCTGAGCAGGTGTTCAATCAAGCAGCGTGGCGAGATGTTGCC 378  
QY 379 GCGGCGCGCAGCTCAAGCTTACCCATGCGGAGCGCACCGTGTTCGCGAGGAGAGCGC 438  
DB 379 AATGGGCGCAGCGAAGCCCAACAGTTGAGGTGATGATCTACGCGCGAAGGCAAGCGGA 438  
QY 439 AGCCATCTCGCGGTCAGATACCGCATCGGCGCGCGCTCGCGCGCTCTGTTGCTGGAG 498  
DB 439 ACGGATTTCTCAGCAGCAGATTCGCGTTCGAGCGCGCTGATGAACTGCTGGAAA 498  
QY 499 CACATCCAGCCATTGTGAAATACGCATGTACGCGCGCAAGCAAGTCCACGTCGCG 558  
DB 499 CATTTCCAAACCGCTCAGCAAGTTTCATGATGTACAGCTCGGTGAGCAGGTCCAGTTGCA 558  
QY 559 TCGGCGCGGAGCTTCAAGCTCTATTCGCGGCGCATGCGCTCGGATCGGACCGAGGTCAAT 618  
DB 559 TCGGCGCGGAGATGTCCTCTTCAGCGCGGATTTTCCAACTGAGCATGAAAGCCAAAC 618  
QY 619 ACCGCGCGAAGCATCTACGCGGTGAGGCGCGCTGCTACGTCGTCGTCGTCGCG 678  
DB 619 GCGACGCTCACCGCGCTGTAAGCAATGAAAGCCAAACCTTTGTGCTTCTGACGCGAG 678  
QY 679 ACCGTTTCGCGGAGATGATCAAGGTATGAGTGAATACGCCCGCAAGAGAGATGTTCTC 738  
DB 679 GTGATCGGACCTTAGCGCGCATGAAACGTTCTGCTCAACGACGAAACGCGCATGTTG 738  
QY 739 AAGCGCGCGCGGCTTTTGCATGATTTTGGGCGCGAGCGCGCGCTCGCGCGAGCGG 798  
DB 739 CCGCAGAGATGTGCTGCGCGCGCATTTTACGCCCGGATGAAAGCGAGCTTGGAGACCT 798  
QY 799 CTCGCGGAGACGAAAGGAGATGCTGTCGCGCATATCGACCTCGGATGATCGCTTG 858  
DB 799 CTGCGCGAAGATGCTGAGGAGATCTTGTACGAGAGATCGATCTGAGAGATTCGCTG 858  
QY 859 GCCAAGCGCGCGCGCATTCGCGCGCGCATATTTACGCGCGCAAGCTAAAGCGGCTG 918  
DB 859 GCGAAGGCTGAAGCGCATTCGCGGTGAGGCACTATTCGCGGCTGACGTCGTGCGTCCAG 918  
QY 919 CTGATCGAGC 929  
DB 919 TTGACCCGCG 929

RESULT 5  
US-09-823-373-13  
; Sequence 13; Application US/09823373  
; Patent No. 6870038

GENERAL INFORMATION:  
; APPLICANT: Chauhan, Sarita  
; APPLICANT: DiCosimo, Robert  
; APPLICANT: Payne, Mark  
; APPLICANT: Gavigan, John  
; APPLICANT: Fallon, Robert  
; TITLE OF INVENTION: Isolation and Expression of a Gene for Nitriase from  
; FILE REFERENCE: BC-1032 US NA  
; CURRENT APPLICATION NUMBER: US/09/823,373  
; PRIOR FILING DATE: 2001-03-30  
; PRIOR APPLICATION NUMBER: 60/193,707  
; NUMBER OF SEQ ID NOS: 32  
; SOFTWARE: Microsoft Office 97  
; SEQ ID NO 13  
; LENGTH: 1110  
; TYPE: DNA  
; ORGANISM: Acidovorax facilis  
US-09-823-373-13

Query Match 27.3%; Score 283.8; DB 3; Length 1110;  
Best Local Similarity 57.0%; Pred. No. 2.8e-54;  
Matches 519; Conservative 0; Mismatches 392; Indels 0; Gaps 0;

QY 19 AAGTATCGCGCGCGCGGTCAGAGCCGCGCGGTGTTCTCGATCTTCAGCCGACAGTC 78  
DB 19 AAGTTCCTCGCGGCAACCGTTCAAGGCAAGCGCGGTATGCTCGACGCAAGCAACATC 78  
QY 79 GAGAAAGCATCGCGCTGATGAGAGCGCGCCAGAGCAAGCGTGCCTGATCGCATTC 138  
DB 79 GACAAAGTCGATCGGATCATCGAAGAAAGCTCCCAAAAGGCGCGAGTCTGATCGCTTC 138  
QY 139 CCAGAGACTTGGATTCCCGGCTATCCCTTTGATATGAGTGGCGGCGCGCTTGGGGC 198  
DB 139 CCGAAGATTCATTTCCGGGCTACCCCTATTGGCGGTGCTCGCGACGTAAAGTACAGC 198  
QY 199 ATGCGCTTCGTCAGCGCTATTTGAGAAATTCGCTGTCGCGCGAGCAAGCATGCGAG 258  
DB 199 CTAGCTTTACTTACGCTATCAGAGAAATTCGTTGAGTGAAGTGAACCGTATCGGT 258  
QY 259 GCCCTGCGGATGGCGCCCGCGCCAGCGCATGATGCTGGCGCGCTATAGCGAGCGC 318  
DB 259 CGCTTCAGCTGGCGCGCGCGCGCAACAAATCGCACTGCTATGAGGCTATTCGAGCGG 318  
QY 319 GCGGCGCGCAGCGCTCTATATGAGCGCAGCGCATCTTCGCGCCCGATGGCGATCTGATCGCC 378  
DB 319 GAAGCGGATCGCGCTATCTGAGCAGGTGTTCAATCAAGCAGCGTGGCGAGATGTTGCC 378  
QY 379 GCGGCGCGAAGCTCAAGCTTACCCATGCGGCGCACCGCATGTCGTGCGCGCTATAGCGAGCGC 438  
DB 379 AATCGCGCAGATGTAAGGCCACACACGTTGAGCGTACGATCTACGCGCAAGCAACGGA 438  
QY 439 AGCCATCTCGCGGTGACGATACCGCATCGGCGCGCTTCGCGCGCTCTGTTGTCGAG 498  
DB 439 ACCGATTTCTCAGCAGCAGCTTCGCGGTTCGAGCGCGGTGAGATTGAATGCTGGAA 498  
QY 499 CACATCCAGCAATTCGAAATACGCGCATGTACGCGCGCGAAGCAAGTCCACGTCGCG 558  
DB 499 CATTTCCAAACCGCTCAGCAAGTTTCATGATGATGAGCTGATGAGAGGATCCACGTTGA 558  
QY 559 TCGTGGCGGAGCTTCAGCGCTCTATTCGCGGATGCGCTATGCGCTCGGACCGAGGTCAAT 618  
DB 559 TCGTGGCGGAGATGTCCTCTTCAGCGCGGATGTTTCCAACTGAGCATGAAAGCCAAAC 618  
QY 619 ACCGCGCGAAGCATCTACGCGGTGAGGCGCGCTGCTAAGTGTGTCGTCGCGCG 678  
DB 619 GCGACGCTACCGCGCTGTAAGCAATGAAAGCCAAACCTTTGTGCTTTCGACGAGC 678  
QY 679 ACGTTTCGCGGAGATGATCAAGGTATGAGTGAATACGCCCGCAAGAGAGATGTTCTC 738  
DB 679 GTGATCGGACCTTAGCGCGCATGAAACGTTCTGCTCAACGAGACGAGCGCATGTTG 738

Qy 739 AAGCCGCGGCGGTTTGGCATGATTTTCGGGCCGAGCGCGGCCCGGAGCG 798  
Db 739 CCGAAGGATGTGGCTGGCGCGCATTTACGCCCGGATGAGAGAGCTTGCGAGCT 798  
Qy 799 CTCGCCGAGACCGAAGAGGAGCTGCTGCTGCCCGGATTCGACCTCGGATATCGCTTG 858  
Db 799 CTGGCGAAGATGTGCGAGGGGATCTTGTACGAGAGATCGATCTGGAGCAATCTGCTG 858  
Qy 859 GCCAAGCGCGCGCGGATTCGGCGGGCCACTTTCAGCGGCCGACGTAAAGCGCTGCTG 918  
Db 859 GCGAAGGCTGAGCGGATCCGGTCGGGCACTTTCGCGCTGACGTGCTGCTCGCTCAG 918  
Qy 919 CTGATGACG 929  
Db 919 TTGACCCGCG 929

## RESULT 6

US-09-823-373-15  
; Sequence 15, Application US/09823373  
; Patent No. 6870038  
; GENERAL INFORMATION:  
; APPLICANT: Chauhan, Sarita  
; APPLICANT: DiCosimo, Robert  
; APPLICANT: Payne, Mark  
; APPLICANT: Gavigan, John  
; APPLICANT: Fallon, Robert  
; TITLE OF INVENTION: Isolation and Expression of a Gene for Nitriase from  
; FILE REFERENCE: BC-1032 US NA  
; CURRENT APPLICATION NUMBER: US/09/823,373  
; CURRENT FILING DATE: 2001-03-30  
; PRIOR APPLICATION NUMBER: 60/193,707  
; NUMBER OF SEQ ID NOS: 32  
; SOFTWARE: Microsoft Office 97  
; SEQ ID NO 15  
; LENGTH: 1776  
; TYPE: DNA  
; ORGANISM: Acidovorax delafieldii  
US-09-823-373-15

Query Match 27.3%; Score 283.8; DB 3; Length 1776;

Best Local Similarity 57.0%; Pred. No. 3e-54; Mismatches 392; Indels 0; Gaps 0;

Qy 19 AAGTATCGCGCGCGGCGGATCGAGCGCGGCTGCTCGATCTCGACCGCAAGTC 78  
Db 350 AAGTTCCTCGCGGCAACCGTTCAGGCGAGCGCGGATGCTCGACGCGCAACGATC 409  
Qy 79 GAGAAAGGATCGGCGCTGATCGAGCGCGCGCAAGAGAGCGTGGCGCTGATCGCATTC 138  
Db 410 GACAAGTGTGATCGGATCATCGAAGAGCTGCCCAAGGCGCGAGTCTGATCGCTTTC 469  
Qy 139 CCAAGACTTGATTCGCGGCTATCCCTTTGGATATGCTGGCGCGCGCTTGCGGCG 198  
Db 470 CCGAAGTATTCATTCGCGGCTACCCCTATTGGCGGTGGCTCGGAGCGTAAAGTACGC 529  
Qy 199 ATGCGCTTGTTCGACGCGCTATTTGAGAAATTCGCTGTCGCGCGGAGCAAGAGGCGAG 258  
Db 530 CTAAGCTTTACTTCAAGCTATTCAGAGAAATTCGTTGAGCTAGGTGACGACCGTATGCGT 589  
Qy 259 GCCCTGGGATGCGCGCGCGCGCGCGCATGATGCTGCGCGCGCTATAGAGCGCG 318  
Db 590 CGCCTTCGAGCTGGCGCGCGCGCGCAAAATTCGACTGTCATGCGGCTATTCGAGCGG 649  
Qy 319 GCGGCGCGGACCTTATATGCGCGAGCGATCTTCGCGCGCGCGGATGAGTCTGATCGCC 378  
Db 650 GAAAGCGGATGCGGCTATCTAGCGAGGTGTTCACTGAGCGAGCGTGGGAGATGTTGCC 709  
Qy 379 GCGCGCGCAAGCTCAAGCTTACCATCGAGGAGCGCACTGTTTCGCGGAGGAGAGCGCG 438  
Db 710 AATGCGCGCAAGCTCAAGCGCCACACACTTAAAGCTTACGATCTACGCGGAGGAGCGGA 769

Qy 439 AGCATCTCGCGGTGACAGATACCGCATCGGCGCGCTCGCGCGCTCTGTTGCTGGAG 498  
Db 770 ACCGATTTCTTACGACGACGACTTTCGCTGGAGCGCTGTTGATTAAGTCTGCGGAA 829  
Qy 499 CACATCCAGCATATGTAAGTAATGCGCATGTACGCGCGCGCAAGAGTCCAGTCCGCG 558  
Db 830 CATTCCAGCGCTCAGCAAGTTTATGATGTACAGCGCTCGGTGAGAGGTCCAGTTTCA 889  
Qy 559 TCGTGGCGGAGCTTCAAGCTTCTATCGCGGATGCGCTTATGCGCTCGAAGGAGTCAAT 618  
Db 890 TCGTGGCGGAGTGTCCCTCTTCAAGCGGATTTTCCAACTGACATGAAAGCAAC 949  
Qy 619 ACCGCGCGAAGCATCTACGCGGTGAGAGGCGCGCTGACGTGCGCTGCTGCGCG 678  
Db 950 GCGAGGTACCGCTGCTGAGCAATGAAAGGCCAACTTTGTGCTTGTGCTGCAACGAG 1009  
Qy 679 ACCGTTTCGCGGAGATCAAGGTATGTTGATGATCGCGCGCAAGAGATGTTCTTC 738  
Db 1010 GTGATCGGACTTACGCGGATCGAAGCTTCTGCTCAAGCAAGAGCGGCACTGTTG 1069  
Qy 739 AAGCCGCGCGCGGTTTTCGATGATTTTCGGGCGCGACGCGCGCGCTGCGGAGCG 798  
Db 1070 CCGAAGATGTGGCTGGCGCGCATTTACGCCCGGATGAGAGCGAGCTTGCGAAGCT 1129  
Qy 799 CTCGCCGAGACCGAAGAGGAGCTGCTGCGCGCATTCGACCTCGGCAATGATCGCTTG 858  
Db 1130 CTGCGGAGAAATGTGAGGGGATCTTGTACGAGATGATTCGAGAGAGATTTCTGCTG 1189  
Qy 859 GCCAAGCGCGCGGATTCGCGCGCGCGCATATTCAGCGCGCGGATGACGCGCTGCTG 918  
Db 1190 GCGAAGCTGAGCGGATTCGGTGGGCACTTATTCGCGGCTGACGTGCTGCTGCTCAG 1249  
Qy 919 CTGATGACG 929  
Db 1250 TTGACCCGCG 1260

## RESULT 7

US-09-823-373-16  
; Sequence 16, Application US/09823373  
; Patent No. 6870038  
; GENERAL INFORMATION:  
; APPLICANT: Chauhan, Sarita  
; APPLICANT: DiCosimo, Robert  
; APPLICANT: Payne, Mark  
; APPLICANT: Gavigan, John  
; APPLICANT: Fallon, Robert  
; TITLE OF INVENTION: Isolation and Expression of a Gene for Nitriase from  
; FILE REFERENCE: BC-1032 US NA  
; CURRENT APPLICATION NUMBER: US/09/823,373  
; CURRENT FILING DATE: 2001-03-30  
; PRIOR APPLICATION NUMBER: 60/193,707  
; NUMBER OF SEQ ID NOS: 32  
; SOFTWARE: Microsoft Office 97  
; SEQ ID NO 16  
; LENGTH: 1110  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: A synthetic  
; OTHER INFORMATION: version of the nitriase gene  
US-09-823-373-16

Query Match 19.4%; Score 202; DB 3; Length 1110;

Best Local Similarity 51.3%; Pred. No. 5.1e-36; Mismatches 445; Indels 0; Gaps 0;

Qy 12 CATGACGAAGTATGCGCGCGCGGCTGAGCGCGCGCGGCTGCTGATCTGACCG 71  
Db 12 CACTCCAAAGTTCTTGCTGCTACTGTTCAAGCTGAGCGAGTTTGTGAGCGAGACGC 71

QY 72 CACAGTGGAGAAAGGATCGGCTGATGAGACGGCGCCAAAGACGATGCGCTTAT 131  
Db 72 TACTATGACAAAGTCTATGATCATCAAGAAAGCTGCCAAAAGGGTGCCTCTTGGAT 131  
QY 132 CGCATTTCCAGAGACTTGATTTCCGAGCTATCCCTTTTGGATATGAGTGGCGCGCGC 191  
Db 132 CGCTTTCCAGAGATTTTCAATCCAGGTTACCACTACGAGGCGCTGGTTGGGTGAAGTTAA 191  
QY 192 TTGGGGGACATGCGCTTCGTCAGCGCTATTTGAGAAATTCGCTCGTGGCGGACAGACA 251  
Db 192 GTACTCTTTGCTTTTACTTCAAGATATCAAGAACTCTTGGAGTTGGGTGAAGACAG 251  
QY 252 GTGGCAGGCGCTTGCGGATGCGCGCGCGCGCGCGCGCGCGCGCTATAG 311  
Db 252 AATGCGTAGACTGCAATTTGGCTGCCCGTAAGAAACAAATTCCTTTGGTACGTTATTC 311  
QY 312 CGAAGCGCGGCGCGAGACCTCTATATGAGGCGAGCGCATCTTTCGCGCCCGATGGCGATCT 371  
Db 312 CGAAGAGAAAGCTGGATCTCGTTACTTGTCCAAAGTCTTCAGCAGAGAGAGGTGAGAT 371  
QY 372 GATGCGCGCGCGCGCGAGCTCAAGCCTACCGCATGCGGAGCGCACCGGTTGCGCGAGGG 431  
Db 372 TGTTCGAATGATGTGTATGTTGAAGCCAACTCAAGTGAAGCATCTACGAGAAAG 431  
QY 432 AGACGCGAGCCATCTCGCGGTGACAGATCCCGCATCGGCGCGCTCGCGCTCTGTG 491  
Db 432 TAAAGGAAACGATTTCTTGACTCAAGACTTCCGCTTGGAGAGAGTTGGTGAATGAACTG 491  
QY 492 CTGGAGACACATCCAGCATTTGTGAAATACGCATATACCGCGCGCGCGAGACAGTCCA 551  
Db 492 TTGGAAACATTTCCAACTCTGTCTAAGTTATGATGATACCTCTTGGGTGAAGAGTCCA 551  
QY 552 CGTGCCTGTGCGCGAGCTTCAAGCTCTATCGCGGACATGAGCTATGCGCTCGAGCGGA 611  
Db 552 CGTTGCTTTCTTGGCAGGTATGTCCTCTTCAACGAGATTTTCCAAATGTTCCATCGA 611  
QY 612 GGTGAATACCGCGCGCAAGCATCTACCGGCTCGAGGCGCGCTGCTACGTCGTGGCGTC 671  
Db 612 AGCCAAAGCGCACCTCTACAGATCTTACGCCATCGAAAGTCAAACTTTTGTCTTTCCTC 671  
QY 672 GTGCGCAGCGTTCGCGGAGATCAAGTATGTTGGATATGCGCCGACAGAGAT 731  
Db 672 TACCAAGTATTTGAACTTCTGCTATGAAACCTTCTGTGTAAGACAGAGAGAGC 731  
QY 732 GTTCTCAAGCGCGCGCGGTTTTCATGATTTTCGCGCGCGCGCGCGCTGAGC 791  
Db 732 TTGTGTCCACAAAGATGTGGTGGCAAGAAATTTACGGTCAAGATGATCTGAGCTTGC 791  
QY 792 CGAGCGCGCTCCCGAGACCGAAGGAGACTGTGTGCGCGATATCGAAGCTCGCATGAT 851  
Db 792 CAAGCCTTTGGCTGAAGATGTGGGTATTTTGTACGCTGAGATTCATTTGAGCAAAAT 851  
QY 852 CGCGTTGCGCAAGCGCGCGCGCATCCGCGGCGCGCATATTCAGCGCGCGCGAGTACGCG 911  
Db 852 TCGGCTGGCCAAAGCTGAGAGCGCATCCAGTGGTCACTACCTCCAGACCTGACGCTTGTG 911  
QY 912 GCTGCTGCTGATC 925  
Db 912 GGTCAAGTTCGACC 925

RESULT 8  
US-09-823-373-3  
Sequence 3, Application US/09823373  
Patent No. 6870038  
GENERAL INFORMATION:  
APPLICANT: Chauhan, Sarita  
APPLICANT: DiCosimo, Robert  
APPLICANT: Payne, Mark  
APPLICANT: Gayagan, John  
APPLICANT: Fallon, Robert  
TITLE OF INVENTION: Isolation and Expression of a Gene for Nitrilase from

FILE OF INVENTION: Acidovorax Facilis 72W  
FILE REFERENCE: BC-1032 US NA  
CURRENT APPLICATION NUMBER: US/09/823, 373  
CURRENT FILING DATE: 2001-03-30  
PRIOR APPLICATION NUMBER: 60/193,707  
PRIOR FILING DATE: 2000-03-31  
NUMBER OF SEQ ID NOS: 32  
SOFTWARE: Microsoft Office 97  
SEQ ID NO 3  
LENGTH: 385  
TYPE: DNA  
ORGANISM: Acidovorax facilis  
US-09-823-373-3

Query Match 12.7%; Score 132.2; DB 3; Length 385;  
Best Local Similarity 59.0%; Pred. No. 1,7e-20;  
Matches 227; Conservative 0; Mismatches 158; Indels 0; Gaps 0;

QY 165 CTTTGTGATATGCGTGGCGCGCGCGGCTTGAGGAGATGCGCTTCGTCAGGCGCTATTTGA 224  
Db 1 CTATGGGCGGTGGCTCGGCGACGTAAGTACAGCTTAAGCTTACTTACAGCTATTCAGCA 60  
QY 225 GAATTCGCTCGTGGCGCGCAAGCAAGTGGCAGGCCCTGCGAGATGCGCGCGCGCA 284  
Db 61 GAATTCGTTGAGAGCTAGTGACGACCGTATGCGTCCGCTCAGCTGGCGCGCGCGCAA 120  
QY 285 CGGATGATGATGCTGTGGCGCGCGCTATAGCGAGCGCGCGGCGGAGAGCTCTATATGAGCCA 344  
Db 121 CAATATGCACTGTCTATGAGGCTATTTGAGCGGAGGAGATCGGCTATCTGAGCCA 180  
QY 345 GCGGATCTTGGGCGCGCGAGTGCATGATCGCGCGCGCGCGCAAGCTCAAGCTTACCA 404  
Db 181 GGTGTCATGACGAGGCGGTGGCGAGATCGTTGCCAATCGCGCGCAAGCTGAAAGCCACACA 240  
QY 405 TGGGAGAGCGACCGTGTGCGCGAGGAGACCGGACGCACTTCGCGGTGACAGATACCGC 464  
Db 241 CGTTGAGCGTACATCTACCGCGAAGCGCAAGCACTTCTTCTCAACGCGTCAAGAGTTCCG 300  
QY 465 CATCGGCGCGCTCGCGCGCTCTGTTGCTGGAGACATCCAGCAGCATTTGCAAAATACGC 524  
Db 301 GTTCGAGCGGCTGGTGGATTTGAATGCTGGAAACATTTTCAACCGCTCAGCAAGTTGAT 360  
QY 525 CATGACCGCGCGCGAGCAAGAGTGC 549  
Db 361 GATGTACAGCCTCGTGGAGCAGGTC 385

RESULT 9  
US-08-447-702-4  
Sequence 4, Application US/08447702  
Patent No. 5629190  
GENERAL INFORMATION:  
APPLICANT: Petre, Dominique  
APPLICANT: Carbeleaud, Edith  
APPLICANT: Levy-Schli, Sophie  
APPLICANT: Crozet, Joel  
TITLE OF INVENTION: Polypeptides Possessing A Nitrilase  
TITLE OF INVENTION: Activity, DNA Sequence Coding for Said Polypeptides,  
Patent No. 5629190  
TITLE OF INVENTION: Expression Cassettes and Host Microorganisms Enabling Them  
TITLE OF INVENTION: to be Obtained, and Method of Converting Nitriles to Carboxy-  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS  
STREET: P.O. Box 1404  
City: Alexandria  
STATE: VA  
COUNTRY: USA  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible



OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/447,702  
FILING DATE: 23-MAY-1995  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/194,588  
FILING DATE: 10-FEB-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: FR 92-09-882  
FILING DATE: 10-AUG-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: McGowan, Malcolm K.  
REGISTRATION NUMBER: 39,300  
REFERENCE/DOCKET NUMBER: 003025-019  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 703-836-6620  
TELEFAX: 703-836-2021  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1194 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 87..1148  
US-08-447-702-4

Query Match 10.5%; Score 109.6; DB 2; Length 1194;  
Best Local Similarity 49.8%; Pred. No. 2.2e-15;  
Matches 277; Conservative 0; Mismatches 279; Indels 0; Gaps 0;

31 GCGGCGGTGACGAGCGCGCGGTCTCTGATCTGACGCGACAGTCGAGAAAGCATC 90  
114 GCGAGATGCAAGCTGCTCTGTATTTATGATCTAGAGCAAGTAGATAAACTTGT 173  
91 GCGCTGATCGACGAGCGCGCCAGACAGACGTCGCGCTGATCGCATTCGCCAGACTTGG 150  
174 AAGTTAATGACAGAGACGATCTATGAGCGCCAGAGTTATCGGCTTCCGAGAACATTT 233  
151 ATTCCCGGCTATCCCTTTTGATATGCTGAGCGCGCGCTTGGGCGATGCGCTTGTCTC 210  
234 ATTCCCGGCTATCCATATGATTTGACATCAATATGACCTTCATGATGATGTGG 293  
211 CAGCGCTATTTGAGATTCGCTGTCGCGCGCGACAGACGTCGAGCGCGCTGCGGAT 270  
294 GCGCTCTTTTCAAGATGCGATTGAATCCAGCAAGAGATTCAACAAATTAGTAT 353  
271 GCGGCGCGCGCGACGCGCATGTCGTGCGCGCTATAGCGAGCGCGCGCGAGC 330  
354 GCTGCAAAAAAGATGAGTTAGTTGCTTCTGTATCAGAGAAAGATTAATGCTCG 413  
331 CTCTAATGAGCGAGCGGATTTGCGCCCGGATGCGATCTGATCGCGCGCGCGCAAG 390  
414 CTATATTTGACGCAATTTGCTTTGACCCGATGATGATTTGATTTGACAGCAAGAAAA 473  
391 CTGAGGCTCCCATGCGAGCGGACGCGTGTGCGGCAAGGAGCGGACGACATCTGGCG 450  
474 TTCAAGCGCCCTAGTAGTAAAGCTGTATGAGGAGATGAGGAGTAGAAGATGGCTCC 533  
451 GTGACAGATACCGCATCGGCGCGCTCGCGCGCTCTGTTGCTGGAGACATTCACAGCA 510  
534 GTATTTAAACAGAGATAGGGAATCTTAGGGGACTTCAGAGTCTGGAAACATGCTCCCA 593  
511 TTGTGCAAAATACGCAATGATAGCGCGCGAGCAAGAGTTCACGTCGCGCTGTGGCCGAGC 570  
594 TTAAACATTTGGGAGATGGGCTCATTTGAACGAGGATACATGTTGCTTCTGGCCAGCC 653  
571 TTCAGCCTTATCGCG 586

Db 654 TTGCTCCCTAAGCG 669

RESULT 10  
US-08-465-615-4  
Sequence 4, Application US/08465615  
Patent No. 5635391  
GENERAL INFORMATION:  
APPLICANT: PETRE, Dominique  
APPLICANT: CERBELEAUD, Edith  
APPLICANT: LEVY-SCHIL, Sophie  
APPLICANT: CROUZET, Joel  
TITLE OF INVENTION: POLYPEPTIDES POSSESSING A NITRILASE  
TITLE OF INVENTION: ACTIVITY, DNA SEQUENCE CODING FOR SAID POLYPEPTIDES  
TITLE OF INVENTION: EXPRESSION CASSETTES AND HOST MICROORGANISMS ENABLING THEM  
TITLE OF INVENTION: TO BE OBTAINED, AND METHOD OF CONVERTING NITRILES TO  
CARBOXYLATES BY MEANS OF SAID POLYPEPTIDE  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/465,615  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/194,588  
FILING DATE: 10-FEB-1994  
APPLICATION NUMBER: FR 9209882  
FILING DATE: 10-AUG-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Crane-Feury, Sharon E  
REGISTRATION NUMBER: 36,113  
REFERENCE/DOCKET NUMBER: 003025-015  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1194 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-465-615-4

Query Match 10.4%; Score 108; DB 2; Length 1194;  
Best Local Similarity 49.8%; Pred. No. 5e-15;  
Matches 276; Conservative 0; Mismatches 280; Indels 0; Gaps 0;

31 GCGGCGGTGAGGCGCGCGCGGTCTCTGATCTGACCGCGACGTCGAGAAAGCATC 90  
114 GCGAGATGCAAGCTGCTCTGTATTTATGATCTAGAGCAAGTAGATAAACTTGT 173  
91 GCGCTGATCGACGAGCGCGCCAGACAGACGTCGCGCTGATCGCATTCGCCAGACTTGG 150  
174 AAGTTAATGACAGAGACGATCTATGAGCGCCAGAGTTATCGGCTTCCGAGAACATTT 233  
151 ATTCCCGGCTATCCCTTTTGATATGCTGAGCGCGCGCTTGGGCGATGCGCTTGTCTC 210  
234 ATTCCCGGCTATCCATATGATTTGACATCAATATGACCTTCATGATGATGTGG 293  
211 CAGCGCTATTTGAGATTCGCTGTCGCGCGCGACAGACGTCGAGCGCGCTGCGGAT 270

Db 294 GCCGCTTTTCAAGATGCGATTGAAATCCCAAGCAAGATTCAACAAATTATGAT 353  
QY 271 GCGGCCCCCGCCACCGCATGATGTCGCGCGCTATAGCAGCGCGCGCGCAGC 330  
Db 354 GCTCAAAAAAAGATGAGTTTACGTTTGCTTTCTGATCAGAGAAATATAGCTCG 413  
QY 331 CTCTATATGAGGCGAGCGCATCTTGGCGCCGATGCGCATGATGCGCGCGCGCAG 390  
Db 414 CTATATTTAGCGCAATTGTTGATCCGATGATGATTTGATTTGACACACAGGAAA 473  
QY 391 CTCAGGCTTACCATGCGAGCGCACCGTGTTCGCGAGGAGACGCGAGCATTCGCG 450  
Db 474 TTTAAGCCCACTAGTATGAAAGAGCTGTATGGGAGATGGGATGGAAGCATGCTCC 533  
QY 451 GTGACGATACCGCATGCGCGCGCTCGCGCGCTCTGTTCTGAGAGCATCAGCA 510  
Db 534 GTATTTTAAACAGATGATGGAATCTTGGGAGACTCCAGTGTGGAACATGCTTCCA 593  
QY 511 TTGTGAAATPACGCATGTACGCGCGCACGACGATCCAGTCGCGTCCGCGCAGC 570  
Db 594 TTAACATGCGCGCATGCGCTCATTTGAACGACGATACATGTTGCTTCCGCGAGCC 653  
QY 571 TTCAGCTCTATCCG 586  
Db 654 TTGCTCCCTAAGCG 669

RESULT 11  
US-09-902-540-4902  
; Sequence 4902, Application US/09902540  
; Patent No. 6833447  
; GENERAL INFORMATION:  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Wiegand, Roger C.  
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof  
; FILE REFERENCE: 38-10(15849)B  
; CURRENT APPLICATION NUMBER: US/09/902,540  
; CURRENT FILING DATE: 2001-07-10  
; PRIOR APPLICATION NUMBER: 60/217,883  
; PRIOR FILING DATE: 2000-07-10  
; NUMBER OF SEQ ID NOS: 16825  
; SEQ ID NO 4902  
; LENGTH: 1995  
; TYPE: DNA  
; ORGANISM: Myxococcus xanthus  
US-09-902-540-4902

Query Match 5.5%; Score 57; DB 3; Length 1995;  
Best Local Similarity 48.6%; Pred. No. 0.0013;  
Matches 258; Conservative 0; Mismatches 255; Indels 18; Gaps 3;  
QY 525 CATGTACGCGCGCGACGAAAGATGCTCCAGTCGCGTCTGCGCCGAGCTTCACTTATCG 584  
Db 129 CGTGCTGCGCGCGACGAGCGGTGCGCATTTGCTCTCCGCGAAGAGTCTTACT 188  
QY 585 CGGATGCGCTATGCGCTCGGACCGAGATCAATACCGCGCAGACGATCTTACGCGGT 644  
Db 189 CGTCATCGGGAAGATTTCTGGAAGCGGAAAGACGTCCGGCGCGAGGCAATCCACCGGG 248  
QY 645 CGAGGCGGCTGCTACTGCTGCGGTGTCGCGACCGCTTTCCGCGAGATGATCAAGT 704  
Db 249 CTACGGTTTCTTGCGAAGACGCGACTTCGCGCGCTTGCAGCAGCGGCGCTGCT 308  
QY 705 ATTGATGATACGCGCGCAAGAGATGTTCTCAAGCGCGCGCGCGCTTTTCCATGAT 764  
Db 309 GTTCATGCGCGCGGAGCGAGCGCATCAGCTGATGAGCAACAAGGTCAAGGAGAGCT 368  
QY 765 TTTTGGGCGC--CGAGCGCGCGCGCTG-----GCCGAGCGCTCCGAGAGACCGA 812  
Db 369 GCGCATGATTTGCGCGCGCGCTGCGCTGCTGCTAGCAGAGCGTCCGACTGATGA 428

QY 813 AGAGGACTGCTGCTGCGCGCATATGACCTTCGCGATGATGCGCTTGGCGAAGCGCGGC 872  
Db 429 TGAAGCGCTGCGCGGTGAGGCGAGCGCATTCGCTTCCGCTGATGATCAAGCGCGGC 488  
QY 873 CGATTCGCGCGCGCGCATATTCAGCGCC-----CGAGTAAACCGCTGCTGCTGATG 926  
Db 489 GGGTGGCGCGCGCGCGCATGCGCATGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 548  
QY 927 ACGTCGCGCGCGCGCATGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 986  
Db 549 GCGCGCGCGCGCGCTGCGAGCGAGCATGCTTTCGGAAGCGCGGAGCTCATCTGAGAA 608  
QY 987 GGGGCAAGCGCGCGCGCTGCGGTGCGGAAAGCGCGCGCGCGCGCA 1037  
Db 609 GCGCTATGACGCGCGCGCATGAGTGTCCAGTCTTTCGCGACGCA 659

RESULT 12  
US-09-902-540-1241/c  
; Sequence 1241, Application US/09902540  
; Patent No. 6833447  
; GENERAL INFORMATION:  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Wiegand, Roger C.  
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof  
; FILE REFERENCE: 38-10(15849)B  
; CURRENT APPLICATION NUMBER: US/09/902,540  
; CURRENT FILING DATE: 2001-07-10  
; PRIOR APPLICATION NUMBER: 60/217,883  
; PRIOR FILING DATE: 2000-07-10  
; NUMBER OF SEQ ID NOS: 16825  
; SEQ ID NO 1241  
; LENGTH: 28493  
; TYPE: DNA  
; ORGANISM: Myxococcus xanthus  
; NAME/KEY: unsure  
; LOCATION: (1) .. (28493)  
; OTHER INFORMATION: unsure at all n locations  
US-09-902-540-1241

Query Match 5.5%; Score 57; DB 3; Length 28493;  
Best Local Similarity 48.6%; Pred. No. 0.002; Length 28493;  
Matches 258; Conservative 0; Mismatches 255; Indels 18; Gaps 3;  
QY 525 CATGTACGCGCGCGACGAAAGATGCTCCAGTCGCGTCTGCGCCGAGCTTCACTTATCG 584  
Db 2496 CGTGCTGCGCGCGACGAGCGGTGCGCATTTGCTCTCCGCGAAGAGTCTTACT 2437  
QY 585 CGGATGCGCTATGCGCTCGGACCGGAGATCAATACCGCGCAGACGATCTTACGCGGT 644  
Db 2436 CGTCATCGGGAAGATTTCTGGAAGCGGAAAGACGTCCGGCGCGAGCGCATCCACCGGG 2377  
QY 645 CGAGGCGGCTGCTACTGCTGCGGTGTCGCGACCGCTTTCCGCGAGATGATCAAGT 704  
Db 2376 CTACGGTTTCTTGCGAAGACGCGACTTCGCGCGCTTGCAGCGCGCGCTGCT 2217  
QY 705 ATTGATGATACGCGCGCAAGAGATGTTCTCAAGCGCGCGCGCGCTTTTCCATGAT 764  
Db 2316 GTTCATGCGCGCGAGCGCGAGCGCATCAGCTGATGAGCAACAAGCTTCAAGCGAGCT 2257  
QY 765 TTTTGGGCGC--CGAGCGCGCGCGCTG-----GCCGAGCGCTCCGAGAGACCGA 812  
Db 2256 GCGCATGATTTGCGCGCGCGGTGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 2197  
QY 813 AGAGGACTGCTGCTGCGCGCATATGACCTTCGCGATGATGCGCTTGGCGAAGCGCGGC 872  
Db 2196 TGAAGCGCTGCGCGGTGAGAGCGAGCGCATGCGCTTCCGCTGATGATGTAAGCGCGGC 2137  
QY 873 CGATTCGCGCGCGCATATTCAGCGCC-----CGAGTAAACCGCGCTGCTGATG 926

Db 2136 GGGTGGCGGCGGCGCGGCGATGCGGTGTGTCAGAGGATGCGAGTTTCGGCGCGCCCT 2077  
Qy 927 ACGTCCGCGCCCAAGCGGTGTCTACCGCTTGATGCGGATTCGAACCGCAAAAGAGAGCA 986  
Db 2076 GCGCGCGCGCGCTTCGAGGAGCAAGTCCCTTCGGAGCGGAGGCTCATCTTGAGAA 2017  
Qy 987 GGGCGACGCGCGCGCTGCGGTGTGAGCGGAAAGCGCGCGCGCGCA 1037  
Db 2016 GCGCGTATCGACGCGCGGCACTGAGAGTCCAGGCTTCGCGAGACGCA 1966

## RESULT 13

US-09-758-759-126  
; Sequence 126, Application US/09758759  
; Patent No. 6861513  
; GENERAL INFORMATION:  
; APPLICANT: Hosted, Thomas J.  
; APPLICANT: Wang, Tim X.  
; APPLICANT: Horan, Ann C.  
; TITLE OF INVENTION: Erythromycin Biosynthetic Genes  
; FILE REFERENCE: ID0983X US  
; CURRENT APPLICATION NUMBER: US/09/758,759  
; CURRENT FILING DATE: 2001-01-11  
; PRIOR APPLICATION NUMBER: US 60/175,751  
; PRIOR FILING DATE: 2000-01-12  
; NUMBER OF SEQ ID NOS: 204  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 126  
; LENGTH: 1728  
; TYPE: DNA  
; ORGANISM: Micromonospora carbonacea  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(1728)  
; OTHER INFORMATION: evbs  
US-09-758-759-126

Query Match 5.3%; Score 55; DB 3; Length 1728;  
Best Local Similarity 45.1%; Pred. No. 0.0036;

Matches 328; Conservative 0; Mismatches 390; Indels 9; Gaps 3;

Qy 315 GCGCGCGGCGGCGAGCCTCTATATGCGCGAGGATCTTCGCGCCCGATGCGATCTGAT 374  
Db 210 GCGCGCGGCGCCCAAGGATGAACTGCGGTGTACCAAGCGCGCGCGCTGT 269  
Qy 375 GCGCGCGGCGCAAGCTCAAGCTTACCATTGCGAGCGCACCGTTCGCGAGGAGA 434  
Db 270 GCGCTGGCTCGCGCCAGAGCGCACCGGCGCGGTGTATCGGTATCGACGCGCGCA 329  
Qy 435 GCGAGCCATCTCGCGGTGACGATACCGCATCGGCGCGCTCGCGCGCTCTGTGCTG 494  
Db 330 GCGCTCGGAGATTCGCGGAGCGCACCGCCAGGTGCGACCGCGCGCGCGCGCC 389  
Qy 495 GGAGCAAT---CGAGCATTTGTGAATAAGCCATGAGCGCGCGCGCAAGAGTCCA 551  
Db 390 GCTGCTGTGCGCGCGCGCGCTGCGACCGCGCGGTGCGTACGCGGTGCGGAGCTGA 449  
Qy 552 CGTGCCTGTGCGCGAGCTTCAAGCTTATCGCGGACGAGCTTATGCTTCGAGACCGA 611  
Db 450 GCGCGCGCGCGCGGTATGATGACCGCAGCACAACCGCGCGCGCAAGAGGTACAA 509  
Qy 612 GGTCAATACCGCGCGCAAGCTTATGCGGTGAGGCGCGGTGCTACGCTGCGGCTC 671  
Db 510 GGTTCATCTCGGCGCGAGCTTCGCGGCGAGCTGCGCGCGCGCGCGAGTGTGCGCC 569  
Qy 672 GTGCGGACCGTTTCGCGGAGATGATCAAGTATTTGGTGTGATACGCGCGCAAGAGAT 731  
Db 570 GCGCGACACCGGATCGAGCGCGCGATCGGCGGTGCGCGCGCGCTGCGCGATCGCT 629  
Qy 732 GTTCC---TCAAGCGCGCGCGGTTCGATGATTTTCGCGCGCGAGCGCGCGCT 788  
Db 630 GGGCGCGCGCGGCGGAGGTGTGTGCGGAGAGAGTGTGTGTGTGTGTGTGTGTGTGT 689

Qy 789 GCGGAGCGCGCTCCCGAGACCGAAGAGAGACTGTGTGTCGCGATATCGACT---CGG 845  
Db 690 CGCGGTGTGACCGCGCGGCGCGCGCGCGCGAGCTTGAAGTGTGCTTACAGCGCGCTGACGG 749  
Qy 846 CATGATCGCGTTCGCAAGCGCGCGCGATCGCGCGCGCGCATTTACAGCGCGCGAGT 905  
Db 750 CGTGGGCGCGCGCGGTGTGACCGCGCGCTTTCGCGCGCGCGCGCTTCGCGATCCCGGCGT 809  
Qy 906 AACCGGCTGTGCTGTGATCGACGTCGCGCGCAACGCGTGTGACGCTTATGCGCATTT 965  
Db 810 GGTGCGGAGCGCGGCGGTGCGGACCGGACTTCGACCGTCAAGCTTCCCAACCGCGA 869  
Qy 966 CGAACCGCAAAACGAGGACAAAGGCGACGCGCGCGCGCGGTGCGGAGGAAAGCGC 1025  
Db 870 GGAAGCGGCGCGGTGAGCTCTCTGTGCGCTTCGCGAGCGCACCGGAGCGAGCTGGC 929  
Qy 1026 CGCGCGC 1032  
Db 930 GATCGCC 936

## RESULT 14

US-09-758-759-1  
; Sequence 1, Application US/09758759  
; Patent No. 6861513  
; GENERAL INFORMATION:  
; APPLICANT: Hosted, Thomas J.  
; APPLICANT: Wang, Tim X.  
; APPLICANT: Horan, Ann C.  
; TITLE OF INVENTION: Erythromycin Biosynthetic Genes  
; FILE REFERENCE: ID0983X US  
; CURRENT APPLICATION NUMBER: US/09/758,759  
; CURRENT FILING DATE: 2001-01-11  
; PRIOR APPLICATION NUMBER: US 60/175,751  
; PRIOR FILING DATE: 2000-01-12  
; NUMBER OF SEQ ID NOS: 204  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 1  
; LENGTH: 109519  
; TYPE: DNA  
; ORGANISM: Micromonospora carbonacea  
US-09-758-759-1

Query Match 5.3%; Score 55; DB 3; Length 109519;  
Best Local Similarity 45.1%; Pred. No. 0.0067;

Matches 328; Conservative 0; Mismatches 390; Indels 9; Gaps 3;

Qy 315 GCGCGCGGCGGCGAGCCTCTATATGCGCGAGGATCTTCGCGCCCGATGCGATCTGAT 374  
Db 79000 GCGCGCGGCGCCCAAGGATGAACTGCGCGGTGTACCAAGCGCGCGCGCGCTGT 79059  
Qy 375 GCGCGCGCGCGCAAGCTCAAGCTTACCATTGCGAGCGCACCGTTCGCGAGGAGA 434  
Db 79060 CGCTGTGCTCGCGCGCCAGAGACGCGCACCGGCGCGGTGTATCGGTATCGACGCGCGCA 79119  
Qy 435 GCGAGCCATCTCGCGGTGACGATACCGCATCGGCGCGCTCGCGCGCTCTGTGCTG 494  
Db 79120 GCGCTCGGAGATTCGCGGAGCGCACCGCGCGAGGTGCGCACCGCGCGCGCGCGCC 79179  
Qy 495 GGAGCAAT---CGAGCATTTGTGAATAAGCCATGAGCGCGCGCGCAAGAGTCCA 551  
Db 79180 GCTGCTGTGCGCGCGCGCGCTGCGACCGCGCGGTGCGTACGCGGTGCGGAGCTGA 79239  
Qy 552 CGTGCCTGTGCGCGAGCTTCAAGCTTATCGCGGACGAGCTTATGCTTCGAGACCGA 611  
Db 79240 GCGCGCGCGCGGTATGATGACCGCAGCACAACCGCGCGCGCAAGAGGTACAA 79299  
Qy 612 GGTCAATACCGCGCGCAAGCTTATGCGGTGAGGCGCGGTGCTACGCTGCGGCTC 671  
Db 79300 GGTTCATCTCGGCGCGAGCTTCGCGGCGAGCTGCGGCGCGGCGCGAGTGTGCGCC 79359  
Qy 672 GTGCGGACCGTTTCGCGGAGATGATCAAGTATTTGGTGTGATACGCGCGCAAGAGAT 731

Db 79360 GGCAGACACCGGATCGAAGGCGCCGATCCGGGCGGTGCGCCCGCTGAGCCGACGTAACCGCT 79419  
QY 732 GTTCC--TCAAGCCGCGCGCGGCTTTTGGCATATTTTGGAGCCCGACCGCGCGCTT 788  
Db 79420 GGGGCCCCGGGAGGTCTGCGGACGAGGTGTGTGTCTGATCTGACACCGGCGCG 79479  
QY 789 GGCCGAGCCGCTCCGAGAGACCGAAGAGGAGTCTGTGTCGCGATTTCAAGCT---CGG 845  
Db 79480 CGCGGTGTGAGACCCGCGCGGGGCGCCGAGAGCTGAAGGTGGCTTCAACGCGCGTGCACGG 79539  
QY 846 CATGATGCGCTTGGCCCAAGCGCGCGCGCATCCGCGGCGCACTATTCAAGCGCGACGT 905  
Db 79540 CGTGGGCGCGCGCTGTGTGACCGCGCGCTTTCGCGCGCGCGCTTCGACATCCCGCGGT 79599  
QY 906 AAGCGGCTGCTGTGATGAGTCCGGCCCAAGGAGTGTGTACGCTTATGTCGCGATT 965  
Db 79600 GTGTCCGAGAGAGCGGTGCTCGAGCCGAGACTTCCGACGCTCAAGCTTCCCAACCGGAG 79659  
QY 966 CGAACCGCAAAACGAGGACAAAGGCGACGCGCCCGCTGCGCTGTGTGCGGAAAGCGC 1025  
Db 79660 GGAAGCCGGGGCGGTGACCTCTCTGTGCGCTGCGGAGGCGACGCGGGGAGACCTGGC 79719  
QY 1026 CGCGCGC 1032  
Db 79720 GATCGCC 79726

RESULT 15  
US-10-152-886-94  
; Sequence 94, Application US/10152886  
; Patent No. 6912470  
; GENERAL INFORMATION:  
; APPLICANT: ECOPIA BIOSCIENCES INC.  
; APPLICANT: Farnet, Chris  
; APPLICANT: Steffa, Alfredo  
; APPLICANT: Zazopoulos, Emmanuel  
; TITLE OF INVENTION: GENES AND PROTEINS INVOLVED IN THE BIOSYNTHESIS OF ENEDINE RING  
; FILE REFERENCE: 3011-305  
; CURRENT APPLICATION NUMBER: US/10/152,886  
; CURRENT FILING DATE: 2002-05-21  
; NUMBER OF SEQ ID NOS: 102  
; SOFTWARE: PatentIn version 3.0  
; SEQ ID NO 94  
; LENGTH: 5877  
; TYPE: DNA  
; ORGANISM: Streptomyces citricolor  
US-10-152-886-94

Query Match 5.2%; Score 54.6; DB 3; Length 5877;  
Best Local Similarity 42.6%; Pred. No. 0.0053;  
Matches 345; Conservative 0; Mismatches 459; Indels 3; Gaps 1;

QY 229 TCGCTGTGCGCGGACAGCAAGTGGCAGGCTTGGCGGATGCGCGCGCGCACGCGC 288  
Db 4963 TGGTTCGCGGCTTCTGCGCGGACGACGAGTGTGTGCGGACCCCGGACCGGAGCGG 5022  
QY 289 ATGCATGTCTGTGCGCGCTATAGGAGCGCGCGGCGCGAGCTTATATGAGCGCAGCG 348  
Db 5023 ATGATGACCGCATCCAGTCTGTGCTGTGCGGACGCGCTGCGCGGAGGATGAG 5082  
QY 349 ATCTTGGGCGCGCATGAGCGCATGATGCGCGCGCGCAAGCTCAAGCTTACCATGCG 408  
Db 5083 CGGCTGTGCTGTGCGCGCGCGGCGGACGAGACTCGAGTACGTCTTTCGAGCGCGG 5142  
QY 409 GAGCGCACCTGTGCGCGGAGGAGCGGAGCGGAGCGGATCTCGGCTGACGATACCGCATC 468  
Db 5143 GAGCGCTCGGAGGAGCGGAGCACTAGTCTAGGACTCGAGCTCGG---ACCCCTCG 5199  
QY 469 GGGCGCTTGTGCGCGCTCTGTGTGTGAGCAATCCAGCCATTGTGAAATAGCCCATG 528  
Db 5200 GGAACGCTGTGAAACGCTGGAAGGCGCTGCGCTGTGCTCGCTTCCGCAACGCGCGG 5259

QY 529 TAGCGCGCGGACGAAAGAGTCAAGTGTGCGGTGTGAGCGGAGCTTCAAGCCTTATCGCGC 588  
Db 5260 GCGGCGCGGTGTGCGCGCGGACGATGCTGTGTGATCTGAGCGGCGCTGGAACGGGTA 5319  
QY 589 ATGAGCTTATGCGGTGCGGACCGGAGGTCAATATCCGCGCAAGCATCTTACGCGGTGAG 648  
Db 5320 CTGGGCGGAGCGCGCGGTGTGTGTGAAACCGCGCGCGGACCGCGCACCGGAGCAG 5379  
QY 649 GGGCGCTGTCAAGTGTGCGGTGTGTGCGGACCGCTTTCGCGGAGATGATCAAGATTG 708  
Db 5380 GACCGCGCTTCCGTATCCGAGACCGCGGTGTGCGGCGGTGAGCGCGCGGTGAAAGCTG 5439  
QY 709 GTGATATAGCGCGCGCAAGAGATTTCTTCAAGCGCGCGCGGCTTGTTCATGATTTTC 768  
Db 5440 GCGCACCGGCGGACCGGAGCGCGGAGCGCGGAGCTGAGCGCGCGCGCTGAGAGGACGAG 5499  
QY 769 GGGCGCGAGCGCGCGCTGTGCGGAGCGCGCTCCGAGACCGAAGAGGAGCTGTGTC 828  
Db 5500 GTGTGCGCTTCCGACGACGCGCGGACTGACCTCGCGGTGTGCGCGGAGCGGCTGCGC 5559  
QY 829 GCGGATATGAGCTTCCGCAATGATGCGGTGTGCGGAGCGCGCGCGGATCCGCGGCGGAC 888  
Db 5560 TGCATGTGAGTGTGCGGTCCGAGCGGACCGCGGAGGAGCTGAGCGGCTGTGGGCGC 5619  
QY 889 TATTCAAGCGCGCGGATTAAGCGGCTGTGTGATGACGTCCGCGCGCAACGCGTCTGTC 948  
Db 5620 GTGTGCGCTTCCGCGGAACTGTGTGGAACCGGAGCGGCGGAGCGCGGCGGCTGCGC 5679  
QY 949 AGCTTGTATGCGGCTTTCGAAACCGCAAAACGAGAGCAAGGAGCGCGCGCGCTGCGC 1008  
Db 5680 GGCACGAGGTGTGAGCGCGGTGTGAGTGTGTGCGAAGGCGGTGTGACGACGAGCGG 5739  
QY 1009 GTGTGCGGAAAGCGCGCGCGCGC 1033  
Db 5740 CTGACGCTGAGCGGCTTCAACCGCG 5764

Search completed: April 27, 2006, 01:43:04  
Job time: 216.785 secs



QY 181 GCGCGCGGCTTGGGCGATGCGCTTGGTCAGGCGTATTTGAGAAATTCGCTTGTGCGC 240  
Db 181 GCGCGCGGCTTGGGCGATGCGCTTGGTCAGGCGTATTTGAGAAATTCGCTTGTGCGC 240  
QY 241 GCGAGCAAGCAAGTGGCGAGCGCTTGGCGAATGCGGCGCGCGCGCATGATGATGCGT 300  
Db 241 GCGAGCAAGCAAGTGGCGAGCGCTTGGCGAATGCGGCGCGCGCATGATGATGCGT 300  
QY 301 GCGGCGTATAGCGAGCG 360  
Db 301 GCGGCGTATAGCGAGCG 360  
QY 361 GATGCGCATGATGCG 420  
Db 361 GATGCGCATGATGCG 420  
QY 421 TTCGCGAGGAGAGCGCGAGCGCATTCGCGGTCAGATACCGCGCATCGGCGCGCTGCG 480  
Db 421 TTCGCGAGGAGAGCGCGAGCGCATTCGCGGTCAGATACCGCGCATCGGCGCGCTGCG 480  
QY 481 GCGCTCTGTTGCTGGAGCAGCATTCGCGGATTCGGAATACCGCGCATGTCGCGCGGAC 540  
Db 481 GCGCTCTGTTGCTGGAGCAGCATTCGCGGATTCGGAATACCGCGCATGTCGCGCGGAC 540  
QY 541 GAACAGGTCCACGTCGCGTGGCGCGAGCTTCAGCTTATTCGCGCGCATGCGCTATGCG 600  
Db 541 GAACAGGTCCACGTCGCGTGGCGCGAGCTTCAGCTTATTCGCGCGCATGCGCTATGCG 600  
QY 601 CTGCGAGCGAGGTCATTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 660  
Db 601 CTGCGAGCGAGGTCATTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 660  
QY 661 GTGCTGCGCTGTCG 720  
Db 661 GTGCTGCGCTGTCG 720  
QY 721 GACAGGAGATGTTCTCAAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 780  
Db 721 GACAGGAGATGTTCTCAAGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 780  
QY 781 CCGCGCGCTGTCG 840  
Db 781 CCGCGCGCTGTCG 840  
QY 841 CTGCGAGGATGTCGCGTGGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 900  
Db 841 CTGCGAGGATGTCGCGTGGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 900  
QY 901 GACGTAACGCGGCTGCTGTCGATTCGAGCTTCGCGCGCGCGCGCGCGCGCGCGCG 960  
Db 901 GACGTAACGCGGCTGCTGTCGATTCGAGCTTCGCGCGCGCGCGCGCGCGCGCGCG 960  
QY 961 GCATTCGAAACCGCAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020  
Db 961 GCATTCGAAACCGCAAAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1020  
QY 1021 AGCGGAA 1041  
Db 1021 AGCGGAA 1041

RESULT 2  
US-10-146-772-383  
; Sequence 383, Application US/10146772  
; Publication No. US20030124698A1  
; GENERAL INFORMATION:

; APPLICANT: Short, Jay  
; APPLICANT: Weiner, David  
; APPLICANT: Chaplin, Jennifer  
; APPLICANT: Chi, Ellen  
; APPLICANT: Milan, Aileen  
; APPLICANT: Desantis, Grace  
; APPLICANT: Madden, Mark

; APPLICANT: Burk, Mark  
; TITLE OR INVENTION: Nitrilases  
; FILE REFERENCE: Docket No. US20030124698A1 DIV-013US  
; CURRENT APPLICATION NUMBER: US/10/146,772  
; PRIOR FILING DATE: 2002-05-15  
; PRIOR APPLICATION NUMBER: US 60/309,006  
; PRIOR FILING DATE: 2001-07-30  
; PRIOR APPLICATION NUMBER: US 60/351,336  
; PRIOR FILING DATE: 2002-01-22  
; PRIOR APPLICATION NUMBER: US 60/300,189  
; PRIOR FILING DATE: 2001-06-21  
; PRIOR APPLICATION NUMBER: US 09/751,299  
; PRIOR FILING DATE: 2000-12-28  
; PRIOR APPLICATION NUMBER: US 60/254,414  
; PRIOR FILING DATE: 2000-12-07  
; PRIOR APPLICATION NUMBER: US 60/173,609  
; PRIOR FILING DATE: 1999-12-29  
; NUMBER OF SEQ ID NOS: 386  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 383  
; LENGTH: 1041  
; TYPE: DNA  
; ORGANISM: Unknown  
; FEATURE:  
; OTHER INFORMATION: Obtained from an environmental sample  
US-10-146-772-383

Query Match 100.0%; Score 1041; DB 6; Length 1041;  
Best Local Similarity 100.0%; Pred. No. 5,8e-276; Indels 0; Gaps 0;  
Matches 1041; Conservative 0; Mismatches 0;

QY 1 ATGTGAGAGCCATGACGAAGTATCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 60  
Db 1 ATGTGAGAGCCATGACGAAGTATCGCGGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 60  
QY 61 GATTCGACCGCAGTCCGAGAAAGGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 120  
Db 61 GATTCGACCGCAGTCCGAGAAAGGATCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 120  
QY 121 GTGCGCGCTGTCGATTCGCGAGAGCTTGGATTCGCGCGCGCGCGCGCGCGCGCGCG 180  
Db 121 GTGCGCGCTGTCGATTCGCGAGAGCTTGGATTCGCGCGCGCGCGCGCGCGCGCGCG 180  
QY 181 GCGCGCGCGCTTGGGCGATGCGCTTCTTCAGCGCTATTTGAGAAATTCGCTGCGCG 240  
Db 181 GCGCGCGCGCTTGGGCGATGCGCTTCTTCAGCGCTATTTGAGAAATTCGCTGCGCG 240  
QY 241 GCGAGCAAGCAAGTGGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 300  
Db 241 GCGAGCAAGCAAGTGGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 300  
QY 301 GCGGCGTATAGCGAGCG 360  
Db 301 GCGGCGTATAGCGAGCG 360  
QY 361 GATGCGCATGATGCG 420  
Db 361 GATGCGCATGATGCG 420  
QY 421 TTCGCGAGGAGAGCGCGAGCGCATTCGCGGTCAGATACCGCGCATCGGCGCGCTGCG 480  
Db 421 TTCGCGAGGAGAGCGCGAGCGCATTCGCGGTCAGATACCGCGCATCGGCGCGCTGCG 480  
QY 481 GCGCTCTGTTGCTGGAGCAGCATTCGCGGATTCGGAATACCGCGCATGTCGCGCGGAC 540  
Db 481 GCGCTCTGTTGCTGGAGCAGCATTCGCGGATTCGGAATACCGCGCATGTCGCGCGGAC 540  
QY 541 GAACAGGTCCACGTCGCGTGGCGCGAGCTTCAGCTTATTCGCGCGCATGCGCTATGCG 600  
Db 541 GAACAGGTCCACGTCGCGTGGCGCGAGCTTCAGCTTATTCGCGCGCATGCGCTATGCG 600  
QY 601 CTGCGAGCGAGGTCATTCG 660  
Db 601 CTGCGAGCGAGGTCATTCG 660

Db 601 CTCGACCGGAGATCAATACCGCCGCAAGCCAGATCTACGGGCTGACAGGCGGCTGTAC 660  
Qy 661 GTGCTGGGTGCTGCGGACCGGTTTCGCGGAGATGATCAAGTATTGTGTGATATACGCC 720  
Db 661 GTGCTGGGTGCTGCGGACCGGTTTCGCGGAGATGATCAAGTATTGTGTGATATACGCC 720  
Qy 721 GACAAGAGATGTTCTTCAGAGCGCGCGGCTTTTCATGATTTTCGAGCCGACGCG 780  
Db 721 GACAAGAGATGTTCTTCAGAGCGCGCGGCTTTTCATGATTTTCGAGCCGACGCG 780  
Qy 781 CGCGCCCTGCGGACCGGCTTCGCGAGACCGAAGAGGACTGTGTGTCGCGATATGAC 840  
Db 781 CGCGCCCTGCGGACCGGCTTCGCGAGACCGAAGAGGACTGTGTGTCGCGATATGAC 840  
Qy 841 CTCGGCATGATCGGTGGCCAAAGCGGCGGCGCATCGGCGGCGCACTATTCACGCGCC 900  
Db 841 CTCGGCATGATCGGTGGCCAAAGCGGCGGCGCATCGGCGGCGCACTATTCACGCGCC 900  
Qy 901 GACGTAAACGCGGCTGCTGATGACGTCGCGCCCAACGCGTCTGTCACGCTTGATGCC 960  
Db 901 GACGTAAACGCGGCTGCTGATGACGTCGCGCCCAACGCGTCTGTCACGCTTGATGCC 960  
Qy 961 GCATTGCAACCGCAAAAGAGACAAAGGCGACGCGCGCTGCGGCTGTGGCGGAA 1020  
Db 961 GCATTGCAACCGCAAAAGAGACAAAGGCGACGCGCGCTGCGGCTGTGGCGGAA 1020  
Qy 1021 AGCGCCGCGCGCGCGCGAGTAG 1041  
Db 1021 AGCGCCGCGCGCGCGAGTAG 1041

RESULT 3  
US-10-241-742-383  
; Sequence 383, Application US/10241742  
; Publication No. US2004002147A1  
; GENERAL INFORMATION:  
; APPLICANT: Short, Jay  
; APPLICANT: Weiner, David  
; APPLICANT: Chaplin, Jennifer  
; APPLICANT: Chi, Ellen  
; APPLICANT: Milan, Aileen  
; APPLICANT: Desantis, Grace  
; APPLICANT: Madden, Mark  
; TITLE OF INVENTION: Nitric Oxides  
; FILE REFERENCE: Docket No. US2004002147A1 DIV-013US  
; CURRENT FILING DATE: 2002-09-09  
; PRIOR APPLICATION NUMBER: US/10/146,772  
; PRIOR FILING DATE: 2002-05-15  
; PRIOR APPLICATION NUMBER: US 60/309,006  
; PRIOR FILING DATE: 2001-07-30  
; PRIOR APPLICATION NUMBER: US 60/351,336  
; PRIOR FILING DATE: 2002-01-22  
; PRIOR APPLICATION NUMBER: US 60/300,189  
; PRIOR FILING DATE: 2001-06-21  
; PRIOR APPLICATION NUMBER: US 09/751,299  
; PRIOR FILING DATE: 2000-12-28  
; PRIOR APPLICATION NUMBER: US 60/254,414  
; PRIOR FILING DATE: 2000-12-07  
; PRIOR APPLICATION NUMBER: US 60/173,609  
; PRIOR FILING DATE: 1999-12-29  
; NUMBER OF SEQ ID NOS: 386  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 383  
; LENGTH: 1041  
; TYPE: DNA  
; ORGANISM: Unknown  
; FEATURE:  
; OTHER INFORMATION: Obtained from an environmental sample  
US-10-241-742-383  
Query Match 100.0%; Score 1041; DB 6; Length 1041;

Best Local Similarity 100.0%; Pred. No. 5, 8e-276;  
Matches 1041; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
Qy 1 ATGTGAGAGCCCATGAGAGATATCGGCGCGCGGCTGACAGCGCGCGGCTTCTC 60  
Db 1 ATGTGAGAGCCCATGAGAGATATCGGCGCGCGGCTGACAGCGCGCGGCTTCTC 60  
Qy 61 GATCTGACCGCAGTGCAGAAAGCATGGGCTGATTCAGAGCGGCGCAAGCAGAC 120  
Db 61 GATCTGACCGCAGTGCAGAAAGCATGGGCTGATTCAGAGCGGCGCAAGCAGAC 120  
Qy 121 GTGCGCTGATTCGATTCGAGAGACTTGGATTCGCGCTATTCGATATGCTG 180  
Db 121 GTGCGCTGATTCGATTCGAGAGACTTGGATTCGCGCTATTCGATATGCTG 180  
Qy 181 GCGCGCGCGCTTGGGCGATGCGCTTGTCCAGCGCTATTCAGAAATTGCTGTGCGC 240  
Db 181 GCGCGCGCGCTTGGGCGATGCGCTTGTCCAGCGCTATTCAGAAATTGCTGTGCGC 240  
Qy 241 GCGAGCAAGAGATGCGAGCGCTTGGGAGATGCGGCGCGCAGATGATGCTG 300  
Db 241 GCGAGCAAGAGATGCGAGCGCTTGGGAGATGCGGCGCGCAGATGATGCTG 300  
Qy 301 GCGGCTATATCGAGCGCGCGCGCGAGCTCTATATGAGGCGAGCGATCTTCGCGCC 360  
Db 301 GCGGCTATATCGAGCGCGCGCGCGAGCGCTCTATATGAGGCGAGCGATCTTCGCGCC 360  
Qy 361 GATGCGATTTGATGCGCGCGCGCGCGAGCTCAAGCTCAAGCGAGCGACCGT 420  
Db 361 GATGCGATTTGATGCGCGCGCGCGCGAGCTCAAGCTCAAGCGAGCGACCGT 420  
Qy 421 TTGCGGAGAGGAGACCGGCAAGCTTCGCGGCTGACATTCGCGCGCGCTGCGC 480  
Db 421 TTGCGGAGAGGAGACCGGCAAGCTTCGCGGCTGACATTCGCGCGCGCTGCGC 480  
Qy 481 GCGCTGTGTGCTGAGAGCATTCAGCGCATTTGCAAAATACGCGCATGTAGCGCGAC 540  
Db 481 GCGCTGTGTGCTGAGAGCATTCAGCGCATTTGCAAAATACGCGCATGTAGCGCGAC 540  
Qy 541 GAACAGGTCCAGCGTGGTGGCGAGCTTACGCTTATGCGGCGATGCGCTATGCG 600  
Db 541 GAACAGGTCCAGCGTGGTGGCGAGCTTACGCTTATGCGGCGATGCGCTATGCG 600  
Qy 601 CTGGAACCGGAGTGCATACCGCGCAAGCAGATCTACCGGCTGAGGCGCGCTGTAC 660  
Db 601 CTGGAACCGGAGTGCATACCGCGCAAGCAGATCTACCGGCTGAGGCGCGCTGTAC 660  
Qy 661 GTGCTGCGTGCATGCGGACCGGTTTCGCGGAGATGATCAAGGATTTGTGATACGCC 720  
Db 661 GTGCTGCGTGCATGCGGACCGGTTTCGCGGAGATGATCAAGGATTTGTGATACGCC 720  
Qy 721 GACAAGAGATGTTCTTCAGAGCGCGGCGGCTTTTCATGATTTTCGAGCCGACGCG 780  
Db 721 GACAAGAGATGTTCTTCAGAGCGCGGCGGCTTTTCATGATTTTCGAGCCGACGCG 780  
Qy 781 CGCGCCCTGCGGACCGGCTTCGCGAGACCGAAGAGGACTGTGTGTCGCGATATGAC 840  
Db 781 CGCGCCCTGCGGACCGGCTTCGCGAGACCGAAGAGGACTGTGTGTCGCGATATGAC 840  
Qy 841 CTCGGCATGATCGGTGGCCAAAGCGGCGGCGCATCGGCGGCGCACTATTCACGCGCC 900  
Db 841 CTCGGCATGATCGGTGGCCAAAGCGGCGGCGCATCGGCGGCGCACTATTCACGCGCC 900  
Qy 901 GACGTAAACGCGGCTGCTGATGACGTCGCGCCCAACGCGTCTGTCACGCTTGATGCC 960  
Db 901 GACGTAAACGCGGCTGCTGATGACGTCGCGCCCAACGCGTCTGTCACGCTTGATGCC 960  
Qy 961 GCATTGCAACCGCAAAAGAGACAAAGGCGACGCGCGCTGCGGCTGTGGCGGAA 1020  
Db 961 GCATTGCAACCGCAAAAGAGACAAAGGCGACGCGCGCTGCGGCTGTGGCGGAA 1020  
Qy 1021 AGCGCCGCGCGCGCGAGTAG 1041  
Db 1021 AGCGCCGCGCGCGCGAGTAG 1041





ORGANISM: Unknown  
FEATURE:  
OTHER INFORMATION: Obtained from an environmental sample  
US-10-440-503-383

Query Match 100.0%; Score 1041; DB 7; Length 1041;  
Best Local Similarity 100.0%; Pred. No. 5.8e-276;  
Matches 1041; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 ATGTGGAGCCCATGACGAAGTATCGCGCGCGGCTGACAGCGCGCGGCTGCTC 60
D 1 ATGTGGAGCCCATGACGAAGTATCGCGCGCGGCTGACAGCGCGCGGCTGCTC 60
QY 61 GATTCGACCGGACAGTGGAAAGCCGATCGGCTGATCGAGCGCGGCGGCGGCGGAC 120
D 61 GATTCGACCGGACAGTGGAAAGCCGATCGGCTGATCGAGCGCGGCGGCGGCGGAC 120
QY 121 GTGGCCCTGATCGATTCGACAGACTTGATTCGCGCTATTCCTTTGGATATGCTG 180
D 121 GTGGCCCTGATCGATTCGACAGACTTGATTCGCGCTATTCCTTTGGATATGCTG 180
QY 181 GGGCGCGCGGCTTGGGCGATCGCTTCTGACAGCGCTATTTGAGAAATTCGCTGCGC 240
D 181 GGGCGCGCGGCTTGGGCGATCGCTTCTGACAGCGCTATTTGAGAAATTCGCTGCGC 240
QY 241 GGGCAGCAGCAGTGGCAGGCGGCTTGGGCGATCGCGCGCGGCGGCGGCGATGCTG 300
D 241 GGGCAGCAGCAGTGGCAGGCGGCTTGGGCGATCGCGCGCGGCGGCGGCGATGCTG 300
QY 301 GCGGCTATATAGCAGCGCGCGGCGGCGGCGGCTATATATGAGCGAGCGATCTTGGCGCC 360
D 301 GCGGCTATATAGCAGCGCGCGGCGGCGGCGGCTATATATGAGCGAGCGATCTTGGCGCC 360
QY 361 GATGCGCATGTGATCGCGCGCGGCGGCGGCGGCTATATGAGCGAGCGAGCGG 420
D 361 GATGCGCATGTGATCGCGCGCGGCGGCGGCGGCTATATGAGCGAGCGAGCGG 420
QY 421 TTGCGCGGAGGAGACGCGGAGCATCTCGCGGTGACAGATACCGGCGATGCGCGCTGCGC 480
D 421 TTGCGCGGAGGAGACGCGGAGCATCTCGCGGTGACAGATACCGGCGATGCGCGCTGCGC 480
QY 481 GCGCTCTGTTGCTGGAGACATCATCGGCTTGAATAAGCGCATATGACCGCGGAC 540
D 481 GCGCTCTGTTGCTGGAGACATCATCGGCTTGAATAAGCGCATATGACCGCGGAC 540
QY 541 GAACAGGTCCAAGTCTGCGGTGCGGCGGCTTCAAGCTTATCGCGGCGGCGCTATGCG 600
D 541 GAACAGGTCCAAGTCTGCGGTGCGGCGGCTTCAAGCTTATCGCGGCGGCGCTATGCG 600
QY 601 CTGCGACCGGAGTCAATACCGCGGCGGCGGCTTCAAGCTTATCGCGGCGGCGCTATG 660
D 601 CTGCGACCGGAGTCAATACCGCGGCGGCGGCTTCAAGCTTATCGCGGCGGCGCTATG 660
QY 661 GTGCTGGGTGCTGCGGCGGCGGCTTTCGCGGAGATCAAGTATGATGATGATACGCGC 720
D 661 GTGCTGGGTGCTGCGGCGGCGGCTTTCGCGGAGATCAAGTATGATGATGATACGCGC 720
QY 721 GACAAGAGATGTTCTCAAGCGCGGCGGCTTTCGCGGAGATCAAGTATGATGATGATG 780
D 721 GACAAGAGATGTTCTCAAGCGCGGCGGCTTTCGCGGAGATCAAGTATGATGATGATG 780
QY 781 GCGCGCTGCGGCGGCGGCTTTCGCGGAGATCAAGTATGATGATGATGATGATGATG 840
D 781 GCGCGCTGCGGCGGCGGCTTTCGCGGAGATCAAGTATGATGATGATGATGATGATG 840
QY 841 CTGCGCATGATCGGCTTTCGCGGAGATCAAGTATGATGATGATGATGATGATGATG 900
D 841 CTGCGCATGATCGGCTTTCGCGGAGATCAAGTATGATGATGATGATGATGATGATG 900
QY 901 GACGTAACGCGGCTGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATG 960
D 901 GACGTAACGCGGCTGCTGCTGATGATGATGATGATGATGATGATGATGATGATGATG 960
```

```
QY 961 GCATTGACCGCAAAAGAGAGACAGGCGGCGGCGGCTGCGGCTGCGGAG 1020
D 961 GCATTGACCGCAAAAGAGAGACAGGCGGCGGCGGCTGCGGCTGCGGAG 1020
QY 1021 AGCGCGCGCGCGCGGCGGAG 1041
D 1021 AGCGCGCGCGCGCGGCGGAG 1041
```

RESULT 6  
US-10-461-925-383  
; Sequence 383, Application US/10461925  
; Publication No. US2004005378A1  
; GENERAL INFORMATION:  
; APPLICANT: Mark J. Burk  
; APPLICANT: Desantis, Grace  
; APPLICANT: Morgan, Brian  
; APPLICANT: Zhu, Zoulin  
; TITLE OF INVENTION: PROCESES FOR MAKING (R) -ETHYL 4-CYANO-3-HYDROXYBUTYRIC ACID  
; FILE REFERENCE: 09010-270001  
; CURRENT APPLICATION NUMBER: US/10/461,925  
; PRIOR FILING DATE: 2003-06-13  
; PRIOR APPLICATION NUMBER: US 60/389,317  
; PRIOR FILING DATE: 2002-06-13  
; PRIOR APPLICATION NUMBER: US 60/392,944  
; PRIOR FILING DATE: 2002-06-28  
; NUMBER OF SEQ ID NOS: 386  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 383  
; LENGTH: 1041  
; TYPE: DNA  
; ORGANISM: Unknown  
; OTHER INFORMATION: Obtained from an environmental sample  
US-10-461-925-383

Query Match 100.0%; Score 1041; DB 7; Length 1041;  
Best Local Similarity 100.0%; Pred. No. 5.8e-276;  
Matches 1041; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 ATGTGGAGCCCATGACGAAGTATCGCGCGCGGCTGACAGCGCGGCTGCTC 60
D 1 ATGTGGAGCCCATGACGAAGTATCGCGCGCGGCTGACAGCGCGGCTGCTC 60
QY 61 GATTCGACCGGACAGTGGAAAGCCGATCGGCTGATCGAGCGCGGCGGCGGCGGAC 120
D 61 GATTCGACCGGACAGTGGAAAGCCGATCGGCTGATCGAGCGCGGCGGCGGCGGAC 120
QY 121 GTGGCCCTGATCGATTCGACAGACTTGATTCGCGCTATTCCTTTGGATATGCTG 180
D 121 GTGGCCCTGATCGATTCGACAGACTTGATTCGCGCTATTCCTTTGGATATGCTG 180
QY 181 GGGCGCGCGGCTTGGGCGATCGCTTCTGACAGCGCTATTTGAGAAATTCGCTGCGC 240
D 181 GGGCGCGCGGCTTGGGCGATCGCTTCTGACAGCGCTATTTGAGAAATTCGCTGCGC 240
QY 241 GGGCAGCAGCAGTGGCAGGCGGCTTGGGCGATCGCGCGGCGGCGGCGATGCTG 300
D 241 GGGCAGCAGCAGTGGCAGGCGGCTTGGGCGATCGCGCGGCGGCGGCGATGCTG 300
QY 301 GCGGCTATATAGCAGCGCGGCGGCGGCGGCTATATATGAGCGAGCGATCTTGGCGCC 360
D 301 GCGGCTATATAGCAGCGCGGCGGCGGCGGCTATATATGAGCGAGCGATCTTGGCGCC 360
QY 361 GATGCGCATGTGATCGCGCGCGGCGGCGGCGGCTATATGAGCGAGCGGCGGCGG 420
D 361 GATGCGCATGTGATCGCGCGCGGCGGCGGCGGCTATATGAGCGAGCGGCGGCGG 420
QY 421 TTGCGCGGAGGAGACGCGGAGCATCTCGCGGTGACAGATACCGGCGATGCGGCTGCGC 480
D 421 TTGCGCGGAGGAGACGCGGAGCATCTCGCGGTGACAGATACCGGCGATGCGGCTGCGC 480
QY 481 GCGCTCTGTTGCTGGAGACATCATCGGCTTGAATAAGCGCATATGATGATGATG 540
```

Db 481 GCGCTCTGTTCTGAGGACATCCAGCATTTGTCGAAATTCGCCCATTTACGCCGCCGAC 540  
Qy 541 GAACAGGTCCAGTCGTGCGCTGTCGCGGAGCTTCAAGCTTCAATCGCGGATGCTTATGCG 600  
Db 541 GAACAGGTCCAGTCGTGCGCTGTCGCGGAGCTTCAAGCTTCAATCGCGGATGCTTATGCG 600  
Qy 601 CTCGAGCCGAGGATCAATACCGCGCAAGCAGATCTTACCGGTCGAGGCGGCTGCTAC 660  
Db 601 CTCGAGCCGAGGATCAATACCGCGCAAGCAGATCTTACCGGTCGAGGCGGCTGCTAC 660  
Qy 661 GTGCTGCGCTGTCGCGGACCGCTTTCGCGGAGATGATCAAGATTTGTTGATACGCC 720  
Db 661 GTGCTGCGCTGTCGCGGACCGCTTTCGCGGAGATGATCAAGATTTGTTGATACGCC 720  
Qy 721 GACAAGAGATGTTCTCAAGCCGCGCGCGCTTTTTCGATGATTTTCGCGCCGACGCC 780  
Db 721 GACAAGAGATGTTCTCAAGCCGCGCGCGCTTTTTCGATGATTTTCGCGCCGACGCC 780  
Qy 781 CGCGCCCTGCGCGAGCCGCTCCGAGAGACCGAAGAGGAGCTGTCGTCGCGCATATGAC 840  
Db 781 CGCGCCCTGCGCGAGCCGCTCCGAGAGACCGAAGAGGAGCTGTCGTCGCGCATATGAC 840  
Qy 841 CTCGCGCATGATCGCTTGCGCAAGCGCGCGCGATCCGCGCGGCACTATTCACGCGCC 900  
Db 841 CTCGCGCATGATCGCTTGCGCAAGCGCGCGCGATCCGCGCGGCACTATTCACGCGCC 900  
Qy 901 GACGTAAAGCGGCTGCTGTCGATCGATCGTCGCGCCCAAGCGCTGTCAGCTTATGATGCC 960  
Db 901 GACGTAAAGCGGCTGCTGTCGATCGATCGTCGCGCCCAAGCGCTGTCAGCTTATGATGCC 960  
Qy 961 GCATTGCAACCGCAAAAGAGAGACAGAGCGCGCGCGCTGCGCGTGTGTGCGGAA 1020  
Db 961 GCATTGCAACCGCAAAAGAGAGACAGAGCGCGCGCGCTGCGCGTGTGTGCGGAA 1020  
Qy 1021 AGCGCGCGCGCGCGCGAGTAG 1041  
Db 1021 AGCGCGCGCGCGCGCGAGTAG 1041

RESULT 7  
US-10-146-772-33  
Sequence 33: Application US/10146772  
Publication No. US20030124698A1  
GENERAL INFORMATION:  
APPLICANT: Short, Jay  
APPLICANT: Weiner, David  
APPLICANT: Chaplin, Jennifer  
APPLICANT: Chl, Ellen  
APPLICANT: Milan, Aileen  
APPLICANT: Desantis, Grace  
APPLICANT: Madden, Mark  
TITLE OF INVENTION: Nitrlases  
FILE REFERENCE: Docket No. US20030124698A1 DIV-013US  
CURRENT APPLICATION NUMBER: US/10/146, 772  
PRIOR FILING DATE: 2002-05-15  
PRIOR APPLICATION NUMBER: US 60/309, 006  
PRIOR FILING DATE: 2001-07-30  
PRIOR APPLICATION NUMBER: US 60/351, 336  
PRIOR FILING DATE: 2002-01-22  
PRIOR APPLICATION NUMBER: US 60/300, 189  
PRIOR FILING DATE: 2001-06-21  
PRIOR APPLICATION NUMBER: US 09/751, 299  
PRIOR FILING DATE: 2000-12-28  
PRIOR APPLICATION NUMBER: US 60/254, 414  
PRIOR FILING DATE: 2000-12-07  
PRIOR APPLICATION NUMBER: US 60/173, 609  
PRIOR FILING DATE: 1999-12-29  
NUMBER OF SEQ ID NOS: 386  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 33  
LENGTH: 1026

TYPE: DNA  
ORGANISM: Unknown  
FEATURE:  
OTHER INFORMATION: Obtained from an environmental sample  
US-10-146-772-33  
Query Match 65.5%; Score 682; DB 6; Length 1026;  
Best Local Similarity 79.4%; Pred. No. 2.5e-177;  
Matches 808; Conservative 0; Mismatches 210; Indels 0; Gaps 0;  
Qy 1 ATGTGAGGCCATGACGAAGATTCGCGCGCGCGGTCGAGGCGCGGCTGTTCTC 60  
Db 1 ATGTTAAGTCCCGTGAAGCATATCGCGCGCGCGGTCGAGGCGCGGCTGTTCTC 60  
Qy 61 GATCTGACCGGACAGTTCGAGAAAGGATGCGCTGATCGACAGCGCGGCAAGCAGAC 120  
Db 61 GATCTGACCGGACAGTTCGAGAAAGGATGCGCATATCGACAGCGCGGCAAGCAGAT 120  
Qy 121 GTGCGCTGATTCGATTCGAGAGACTTGATTCGCGGCTATCCCTTTGATATGAGCTG 180  
Db 121 GTGCGCTGATTCGCGCTTTCGGAACCTGATTCGCGCTGATGCTGATGCTGCTC 180  
Qy 181 GCGCGCGCGCTTGGGCGATGCGCTTTCGACGCGCTATTTGAGAAATTCGCTGTCGC 240  
Db 181 GCGCTGCGCGCGCTTGGGCGATGCGCTTTCGACGCGCTATTTGAGAAATTCGCTGTCGC 240  
Qy 241 GCGAGCAAGAGTGGAGCGCGCGCGGATGCGCGCGCGCGCGCGCATGATGCTGTCG 300  
Db 241 GCGAGCAACAGTGAACCGGATGCGCGCGCGCGCGCGCGCGCGCATGACGCTGTC 300  
Qy 301 GCGGCTATAGCAGAGCGCGCGCGCGCGAGCTTATATGAGCGCGAGCATTTTCGCGCC 360  
Db 301 GTCGCGCTTACAGCAGCGCGCGCGGAGGAGCAGCTTATGATGAGCGAGCATTTTCGCGCC 360  
Qy 361 GATGCGGATGATTCGCGCGCGCGCGCGCGAGCTTAAGCTTACCATGCGAGCGCACCGTG 420  
Db 361 GAAAGCGAGCTCATTCGCGCGCGCGCGCGAGCTTAAAGCTTAAAGCAGACCGCGAGCGAGTG 420  
Qy 421 TTGCGCGAGGAGACGAGCGCATTCGCGGATGACAGTACCGCATTCGCGCGCTCGCG 480  
Db 421 TTGCGCGAGGAGGAGCGAGCGCATTCGCGGATGACAGCGGATTCGCGCGCTCGCG 480  
Qy 481 GCGCTCTGTTGCTGGAGGACATCCAGCATTTGTGAAATAGCGCATGTCGCGCGAC 540  
Db 481 GCGCTCTGCTGCGGAGGACATCCAGCGCTTCAAAATAGCGCATGTCGCGCGAAC 540  
Qy 541 GAACAGGTCAATGAGCTGCGGAGCATTCAGCGCGCTTCAAAATAGCGCATGTCGCGCGAAC 540  
Db 541 GAACAGGTCAATGAGCTGCGGAGCATTCAGCGCGCTTCAAAATAGCGCATGTCGCGCGAAC 540  
Qy 601 CTCGAGCCGAGGATCAATACCGCGCAAGCAGATCTACCGGTCGAGGCGGCTGCTAC 660  
Db 601 CTCGAGCCGAGGATCAATACCGCGCAAGCAGATCTACCGGTCGAGGCGGCTGCTAC 660  
Qy 661 GTGCTGCGCTGTCGCGGACCGCTTTCGCGGAGATGATCAAGATTTGTTGATACGCC 720  
Db 661 GTGCTGCGCTGTCGCGGACCGCTTTCGCGGAGATGATCAAGATTTGTTGATACGCC 720  
Qy 721 GACAAGAGATGTTCTCAAGCCGCGCGCGCTTTTTCGATGATTTTCGCGCCGACGCC 780  
Db 721 GACAAGAGATGTTCTCAAGCCGCGCGCGCTTTTTCGATGATTTTCGCGCCGACGCC 780  
Qy 781 CGCGCCCTGCGCGAGCCGCTCCGAGAGACCGAAGAGGAGCTGCTGTCGCGCATATGAC 840  
Db 781 CGCGCCCTGCGCGAGCCGCTCCGAGAGACCGAAGAGGAGCTGCTGTCGCGCATATGAC 840  
Qy 841 CTCGCGCATGATCGCTTGCGCAAGCGCGCGCGATTCGCGCGGCACTATTCACGCGGCC 900  
Db 841 CTCGCGCATGATCGCTTGCGCAAGCGCGCGCGATTCGCGCGGCACTATTCACGCGGCC 900  
Qy 901 GACGTAAAGCGGCTGCTGTCGATCGATCGTCGCGCCCAAGCGCTGTCAGCTTATGATGCC 960  
Db 901 GACGTAAAGCGGCTGCTGTCGATCGATCGTCGCGCCCAAGCGCTGTCAGCTTATGATGCC 960

QY 961 GCATTGCAACCGCAAAACGAGACAAAGGCGCCCGCGCTGCGGTGTGGCCG 1018  
DB 961 CGCTTGAGGTGTGTGACAGAGCGAGCGCGCCGCGCCACGCAACCGCGCAAGCGG 1018

RESULT 8  
US-10-241-742-33

Sequence 33, Application US/10241742  
Publication No. US20040002147A1

GENERAL INFORMATION:

APPLICANT: Short, Jay  
APPLICANT: Weiner, David  
APPLICANT: Chaplin, Jennifer  
APPLICANT: Chi, Ellen  
APPLICANT: Milan, Aileen  
APPLICANT: Desantis, Grace  
APPLICANT: Madden, Mark

APPLICANT: Burk, Mark

TITLE OF INVENTION: Nitrlases

FILE REFERENCE: Docket No. US20040002147A1 DIV-013US

CURRENT FILING DATE: 2002-09-09

PRIOR FILING DATE: 2002-05-15

PRIOR FILING DATE: 2002-05-15

PRIOR FILING DATE: 2001-07-30

PRIOR FILING DATE: 2002-01-22

PRIOR FILING DATE: 2001-06-21

PRIOR FILING DATE: 2000-12-28

PRIOR FILING DATE: 2000-12-28

PRIOR FILING DATE: 2000-12-07

PRIOR FILING DATE: 2000-12-07

PRIOR FILING DATE: 1999-12-29

NUMBER OF SEQ ID NOS: 386

SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 33

LENGTH: 1026

TYPE: DNA

ORGANISM: Unknown

FEATURE:

OTHER INFORMATION: Obtained from an environmental sample

US-10-241-742-33

Query Match 65.5%; Score 682; DB 6; Length 1026;  
Best Local Similarity 79.4%; Pred. No. 2.5e-177; Indels 0; Gaps 0;

Matches 808; Conservative 0; Mismatches 210; Indels 0; Gaps 0;

QY 1 ATGTGAGAGCCCATGCAAGATATCGCGCGCGGTGCAAGCGCGCGGTGTTCTC 60  
DB 1 ATGTAAATCCCGTACCAAGATATCGCGCGCGGTGCAAGCGCGCGGTGTTCTC 60  
QY 61 GATCTGACCGCAAGTGAAGAAAGCATCGGCTGATGAGCAGCGCGCGCAAGAGAC 120  
DB 61 GATCTGACCGCAAGTGAAGAAAGCATCGGCTGATGAGCAGCGCGCGCAAGAGAT 120  
QY 121 GTGGCGCTGATGCAATCCGAGAGCTTGATCCCGGCTATCCCTTTGATATAGGCTG 180  
DB 121 GTGGCGCTGATGCAATCCGAGAGCTTGATCCCGGCTATCCCGCTGATATAGGCTG 180  
QY 181 GGCGCGCGCGCTTGCGGATGCGCTTCTGCAAGCGCTATTTGAGAAATTCGCTGCGC 240  
DB 181 GGCTGCGCGCGCTTGCGGATGCGCTTCTGCAAGCGCTATTTGAGAAATTCGCTGCGC 240  
QY 241 GGCAAGCAAGCGTGAAGCGCGCTTGCGGATGCGCGCGCGCGCAAGCAAGTGTGTC 300  
DB 241 GGCAAGCAAGCGTGAAGCGCGCTTGCGGATGCGCGCGCGCGCAAGCAAGTGTGTC 300  
QY 301 GCCGCTATGAGGAGCGCGCGCGCGCGCGCTATATAGGCGAGCGCATCTTGGCGCC 360  
DB 301 GCCGCTATGAGGAGCGCGCGCGCGCGCGCTATATAGGCGAGCGCATCTTGGCGCC 360

RESULT 9  
US-10-440-523-33

Sequence 33, Application US/10440523

Publication No. US20040014195A1

GENERAL INFORMATION:

APPLICANT: Short, Jay

APPLICANT: Weiner, David

APPLICANT: Chaplin, Jennifer

APPLICANT: Chi, Ellen

APPLICANT: Milan, Aileen

APPLICANT: Desantis, Grace

APPLICANT: Madden, Mark

APPLICANT: Burk, Mark

TITLE OF INVENTION: Nitrlases

FILE REFERENCE: Docket No. US20040014195A1 DIV-013US

CURRENT FILING DATE: 2003-05-15

PRIOR FILING DATE: 2002-05-15

PRIOR FILING DATE: 2002-05-15

PRIOR FILING DATE: 2001-07-30

PRIOR FILING DATE: 2002-01-22

PRIOR FILING DATE: 2001-06-21

PRIOR FILING DATE: 2001-06-21

PRIOR FILING DATE: 2001-06-21

DB 301 GTGCGCTTACAGCGAGCGCGCGGAGGAGCGAGCGCTTACATGAGGCGAGCGATCTTGGCCCC 360  
QY 361 GATGCGCATCTGATCGCGCGCGCGCGCGCGCAAGCTTCAAGCTTACCATGCGAGCGCAAGCTG 420  
DB 361 GATGCGCATCTGATCG 420  
QY 421 TTGCGGAGGAGAGCGGAGCG 480  
DB 421 TTGCGGAGGAGAGCGGAGCG 480  
QY 481 GCGCTGTGTTGCTGCGGAGCAATCGCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 540  
DB 481 GCGCTGTGTTGCTGCGGAGCAATCGCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 540  
QY 541 GCAAGGTCCAGCTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCG 600  
DB 541 GCAAGGTCCAGCTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCG 600  
QY 601 CTGCGAGCGGAGGCAATACCGCGCGCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 660  
DB 601 CTGCGAGCGGAGGCAATACCGCGCGCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 660  
QY 661 GTGCTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTG 720  
DB 661 GTGCTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTG 720  
QY 721 GACCAAGAGATGTTCTTCAAGCG 780  
DB 721 GACCAAGAGATGTTCTTCAAGCG 780  
QY 781 CG 840  
DB 781 CG 840  
QY 841 CTGCGCATGATCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTG 900  
DB 841 CTGCGCATGATCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTGCGGTG 900  
QY 901 GACGTAACGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 960  
DB 901 GACGTAACGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 960  
QY 961 GCATTGCAACCGCAAAACGAGACAAAGGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1018  
DB 961 CGCTTGAGGTGTGTGACAGAGCGAGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCG 1018



QY 361 GATGCGATCTGATCGCGCGCGCGCAAGCTCAAGCCTAACCCATTCGCGAGCGACCGTG 420  
DB 361 GAAAGCGAGCTATCGCGCGCGCGCAAGCTCAAGCCTAACCCATTCGCGAGCGACCGTG 420  
QY 421 TTCCGCGAGGAG 480  
DB 421 TTCCGCGAGGAG 480  
QY 481 GCGCTCTGTGTGCGAG 540  
DB 481 GCGCTCTGTGTGCGAG 540  
QY 541 GAAAG 600  
DB 541 GAAAG 600  
QY 601 CTCCGAGCGAG 660  
DB 601 CTCCGAGCGAG 660  
QY 661 GTGCTGCGTGTGCGAG 720  
DB 661 GTGCTGCGTGTGCGAG 720  
QY 721 GACAAAG 780  
DB 721 GACAAAG 780  
QY 781 CGCGCCCTGCGAG 840  
DB 781 CGCGCCCTGCGAG 840  
QY 841 CTCCGAG 900  
DB 841 CTCCGAG 900  
QY 901 GACGAG 960  
DB 901 GACGAG 960  
QY 961 GCATTGCAAGCGCAAAACGAG 1018  
DB 961 GCATTGCAAGCGCAAAACGAG 1018

## RESULT 11

US-10-461-925-33  
: Sequence 33, Application US/10461925  
: Publication No. US20040053378A1  
: GENERAL INFORMATION:  
: APPLICANT: Mark J. Burk  
: APPLICANT: Desantis, Grace  
: APPLICANT: Morgan, Brian  
: APPLICANT: Zhu, Zoulin  
: TITLE OF INVENTION: PROCESSES FOR MAKING (R)-ETHYL 4-CYANO-3-HYDROXYBUTYRIC ACID  
: FILE REFERENCE: 09010-270001  
: CURRENT APPLICATION NUMBER: US/10/461,925  
: PRIOR FILING DATE: 2003-06-13  
: PRIOR APPLICATION NUMBER: US 60/389,317  
: PRIOR FILING DATE: 2002-06-13  
: PRIOR APPLICATION NUMBER: US 60/392,944  
: NUMBER OF SEQ ID NOS: 386  
: SOFTWARE: FastSeq for Windows Version 4.0  
: SEQ ID NO 33  
: LENGTH: 1026  
: TYPE: DNA  
: ORGANISM: Unknown  
: FEATURE:  
: OTHER INFORMATION: Obtained from an environmental sample  
US-10-461-925-33

Query Match 65.5%; Score 682; DB 7; Length 1026;

Best Local Similarity 79.4%; Pred. No. 2,56-177;  
Matches 808; Conservative 0; Mismatches 210; Indels 0; Gaps 0;

QY 1 ATGTGAGAGCCATGACGAGATATCGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAG 60  
DB 1 ATGTGAGAGCCATGACGAGATATCGCGCGCGCGAGAGAGAGAGAGAGAGAGAGAGAGAG 60  
QY 61 GATCTGAG 120  
DB 61 GATCTGAG 120  
QY 121 GTGCGCTGATCGATTCCTCAGAGATTCGAGATTCCTCCTTTTGGATATGATGCTG 180  
DB 121 GTGCGCTGATCGATTCCTCAGAGATTCGAGATTCCTCCTTTTGGATATGATGCTG 180  
QY 181 GAGCGCGAG 240  
DB 181 GAGCGCGAG 240  
QY 241 GCGAG 300  
DB 241 GCGAG 300  
QY 301 GCGAG 360  
DB 301 GCGAG 360  
QY 361 GATGCGATCTGATCGCGCGCGCGCAAGCTCAAGCCTAACCCATTCGCGAGCGACCGTG 420  
DB 361 GAAAGCGAG 420  
QY 421 TTCCGCGAGGAG 480  
DB 421 TTCCGCGAGGAG 480  
QY 481 GCGCTCTGTGTGCGAG 540  
DB 481 GCGCTCTGTGTGCGAG 540  
QY 541 GAAAG 600  
DB 541 GAAAG 600  
QY 601 CTCCGAG 660  
DB 601 CTCCGAG 660  
QY 661 GTGCTGCGTGTGCGAG 720  
DB 661 GTGCTGCGTGTGCGAG 720  
QY 721 GACAAAG 780  
DB 721 GACAAAG 780  
QY 781 CGCGCCCTGCGAG 840  
DB 781 CGCGCCCTGCGAG 840  
QY 841 CTCCGAG 900  
DB 841 CTCCGAG 900  
QY 901 GACGTAAGCGGCTGTGCTGATCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960  
DB 901 GACGTAAGCGGCTGTGCTGATCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 960  
QY 961 GCATTGCAAGCGCAAAACGAG 1018  
DB 961 GCATTGCAAGCGCAAAACGAG 1018

RESULT 12











```

Db      797  CGGCGCGCTGCGCCCGGACGAGGAGGCTTGGCTTAAGCGACATCGATCTCGGATGA 856
QY      851  TCGCGTTGGCCAGGCGGCGCGGATCGGCGGCGCACTATTCAAGGCGCGAGCTAAAGC 910
Db      857  TTTGGATTGGCCAAAGCGGCGAGCCGATCGGCGGCGCATTAATGCAAGCCCGACGTCAACC 916
QY      911  GGGCTGCTGCTGATCGAAGTCCGGCCCAAGCGGCTGCTCAAGCTTGATCCGCATTGGAAC 970
Db      917  GGGCTTGTGTTCAACATCGGCTGGGTATCGGATCGAGACCATGGCGTTGCCGATCGATG 976
QY      971  CGCAAAACGAGACAAAGGCGAGCGGCGCGCTGCGGCTGCTGCGGAAAGCGCGCCG 1030
Db      977  CGGAGACCAAGGCGGAAAGCAAGGCTAAAGCTGGAACCAAGGCAAGCAAGCGTGGCGCGT 1036
QY      1031  CCGCGCAG 1038
Db      1037  TCGGCGCG 1044

```

Search completed: April 27, 2006, 00:01:33  
 Job time : 1004.99 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 26, 2006, 23:27:34 ; Search time 626.12 Seconds  
(without alignments)  
6752.353 Million cell updates/sec

Title: US-09-751-299-1

Perfect score: 1041  
Sequence: 1 atgtcgcagccatgacgaa.....gcgcgcgcgcgcgcagtag 1041

Scoring table: IDENTITY NUC  
Gap0 10.0 , Gapext 1.0

Searched: 9295968 segs, 2030634719 residues

Total number of hits satisfying chosen parameters: 18591936

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published Applications NA\_New:\*  
1: /SID55/ptocdata/2/pubpna/US08\_NEW\_PUB.seq:\*  
2: /SID55/ptocdata/2/pubpna/US06\_NEW\_PUB.seq:\*  
3: /SID55/ptocdata/2/pubpna/US07\_NEW\_PUB.seq:\*  
4: /SID55/ptocdata/2/pubpna/PCT\_NEW\_PUB.seq:\*  
5: /SID55/ptocdata/2/pubpna/US09\_NEW\_PUB.seq:\*  
6: /SID55/ptocdata/2/pubpna/US10\_NEW\_PUB.seq:\*  
7: /SID55/ptocdata/2/pubpna/US10\_NEW\_PUB.seq:\*  
8: /SID55/ptocdata/2/pubpna/US10\_NEW\_PUB.seq:\*  
9: /SID55/ptocdata/2/pubpna/US10\_NEW\_PUB.seq:\*  
10: /SID55/ptocdata/2/pubpna/US10\_NEW\_PUB.seq:\*  
11: /SID55/ptocdata/2/pubpna/US11\_NEW\_PUB.seq:\*  
12: /SID55/ptocdata/2/pubpna/US11\_NEW\_PUB.seq:\*  
13: /SID55/ptocdata/2/pubpna/US11\_NEW\_PUB.seq:\*  
14: /SID55/ptocdata/2/pubpna/US11\_NEW\_PUB.seq:\*  
15: /SID55/ptocdata/2/pubpna/US60\_NEW\_PUB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	297.4	28.6	1071	9	US-10-537-075-6
2	283.8	27.3	1110	9	US-10-919-182-3
3	283.8	27.3	1110	9	US-10-919-182-7
4	283.8	27.3	1110	9	US-10-919-182-11
5	282.2	27.1	1110	9	US-10-919-182-5
6	282.2	27.1	1110	9	US-10-919-182-13
7	282.2	27.1	1110	9	US-10-919-182-15
8	282.2	27.1	1110	9	US-10-919-182-17
9	50.6	4.9	9975	7	US-10-496-351-4
10	50.6	4.9	82746	7	US-10-496-351-56
11	48.2	4.6	1706	11	US-11-096-568A-20685
12	47.8	4.6	113193	7	US-10-478-943D-1
13	47.6	4.6	2247	7	US-10-496-351-8
14	47.6	4.6	4770	7	US-10-496-351-3
15	47.6	4.6	8192	9	US-10-933-746-36
16	47	4.5	720	13	US-11-079-906-1
17	47	4.5	720	14	US-11-079-476-1
18	47	4.5	726	14	US-11-175-690-115

19	47	4.5	1314	14	US-11-069-642-46	Sequence 46, Appl
20	47	4.5	1314	14	US-11-069-642-48	Sequence 48, Appl
21	47	4.5	1314	14	US-11-069-642-50	Sequence 50, Appl
22	47	4.5	1314	14	US-11-069-642-52	Sequence 52, Appl
23	47	4.5	1314	14	US-11-069-642-54	Sequence 54, Appl
24	47	4.5	1314	14	US-11-069-642-56	Sequence 56, Appl
25	47	4.5	1314	14	US-11-069-642-58	Sequence 58, Appl
26	47	4.5	1314	14	US-11-069-642-60	Sequence 60, Appl
27	47	4.5	1314	14	US-11-069-642-62	Sequence 62, Appl
28	47	4.5	1539	14	US-11-032-236-5	Sequence 5, Appl
29	47	4.5	1781	14	US-11-108-890A-11	Sequence 11, Appl
30	47	4.5	2349	11	US-11-194-991-76	Sequence 76, Appl
31	47	4.5	2355	11	US-11-194-991-78	Sequence 78, Appl
32	47	4.5	2502	9	US-10-537-971-1	Sequence 1, Appl
33	47	4.5	3660	14	US-11-032-236-3	Sequence 3, Appl
34	47	4.5	4151	14	US-11-213-368-15	Sequence 15, Appl
35	47	4.5	4862	14	US-11-082-154A-67	Sequence 87, Appl
36	47	4.5	4894	14	US-11-181-148-2	Sequence 2, Appl
37	47	4.5	5041	14	US-11-193-750-6	Sequence 6, Appl
38	47	4.5	5094	9	US-10-948-344-1	Sequence 1, Appl
39	47	4.5	5162	14	US-11-082-154A-26	Sequence 26, Appl
40	47	4.5	5310	14	US-11-082-154A-71	Sequence 71, Appl
41	47	4.5	6119	14	US-11-082-154A-126	Sequence 126, App
42	47	4.5	6748	8	US-10-655-872-3	Sequence 3, Appl
43	47	4.5	7350	8	US-10-655-872-8	Sequence 8, Appl
44	47	4.5	7487	14	US-11-181-148-4	Sequence 4, Appl
45	47	4.5	7600	14	US-11-082-154A-115	Sequence 115, App

## ALIGNMENTS

RESULT 1  
US-10-537-075-6  
; Sequence 6, Application US/10537075  
; Publication No. US20060014291A1  
; GENERAL INFORMATION:  
; APPLICANT: Kesseler, Maria  
; APPLICANT: Zelnickl, Thomas  
; TITLE OF INVENTION: L-RHAMNOS- INDUCIBLE EXPRESSION SYSTEMS  
; FILE REFERENCE: 12810-00091-US  
; CURRENT APPLICATION NUMBER: US/10/537, 075  
; PRIOR APPLICATION NUMBER: 2005-06-01  
; PRIOR FILING DATE: 2003-11-27  
; PRIOR APPLICATION NUMBER: PCT/EP2003/013367  
; PRIOR FILING DATE: 2002-12-02  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 6  
; LENGTH: 1071  
; TYPE: DNA  
; ORGANISM: Alcaligenes faecalis  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(1068)  
; OTHER INFORMATION: coding for nitrilase  
US-10-537-075-6

Query Match 28.6%; Score 297.4; DB 9; Length 1071;  
Best Local Similarity 58.2%; Pred. No. 2.3e-65;  
Matches 523; Conservative 0; Mismatches 376; Indels 0; Gaps 0;

QY 25 CGCGGCGCGCGTGCAGCGCGCGGTTCCTGATCTGCAGCCGACAGTCGAGAA 84  
DB 22 CGGCGAGCGCGCTACAGCGCGCTCTCCCACTACGATCTGCACAGCGGTGTATAA 81  
QY 85 GCGATGCGCGCTGATCAGAGCGCGCCAGACGAGAGCTGCGCTGATCGCATTCAGAG 144  
DB 82 ACCATGACCTGCTGCTGCTGAGCGCGCCGATGAGGCGCTGACCTGATCGTGTGTGAA 141  
QY 145 ACTTGATTCGCGGCTATCCCTTTTGATATAGCTGCGGCGCGCGCTGGGCGCATGCGC 204

```
Db 142 ACCTGGCTGCCGGATATCCCTTCCACGCTCTGGCTGGCGGACCGGCGCTGGTCCGTGAAA 201
Qy 205 TTCTCCAGCGCTATTTTGAAGATTTCCGTCTGGCGGCGGACCAAGACAGTGGCAGCCCTG 264
Db 202 TACAGTCCCTCTACTATGACCACTCGCTCTGCTGGACAGTGGAGAGTTTCAACGCAAT 261
Qy 265 GCGAGTGGCGCGCGCGGACGAGCATGTGATGCTGGCGGCTATAGAGAGCGGCGCGG 324
Db 262 GCCAGAGCGGACGAGCCTTGGGTATTTTCACTGGCATGGGTATAGGAGCGGACGCG 321
Qy 325 GGCAGCTCTATATATGGCGCAGGCGCATCTTCCGCCCCGATGGCGATCTGATGCCGCGC 384
Db 322 GGCAGCTCTTATCTGGGCGCAATGCTGTATGACGACCAAGGCGAGATGCTGTGGTGG 381
Qy 385 CGCAAGCTCAAGCCTTACCCATGCGGAGCGACCGCTGTTCCGCGAGGAGACGCGACCAT 444
Db 382 CGCAAACTCAAAACCGACGATGTAGAGCGACCGTATTTGGTGAAGGTTATGCCGTGAT 441
Qy 445 CTCCGCGGTGACAGATACCGGCATCGGGGCGCTCGGCGCGCTCTGTGCTGGGAGACATC 504
Db 442 CTGATTTGTCTCGACACAGAACTGGAGCGCTCGGTCTTATCTGCTGGAGCATTTG 501
Qy 505 CAGCATTTGTCAAAATACGCGATGTACGCGCGCGACGAAAGGTCCACGTCGCTGCG 564
Db 502 TCGCCCTTGAGCAAGTACGCGCTGTACTCCAGCATGAAGCATTTACATTTGCTGCTGG 561
Qy 565 CCGAGCTTCAAGCCTCTATTCGCGCATGGGCGTATGGCTCGGACCGGAGGTCAATACCGCC 624
Db 562 CCGTGTCTTTGCTATATACGGAACAGGCGCCCGCTCGGTGCAAGGTGAACATGCT 621
Qy 625 GCAAGCCAGATCTTACGCGGTGAGAGCGGCTGCTACGTGCTGCGTGGCGGACCGTT 684
Db 622 GCTCGCAAACTCTATGCGTTGAAGGCGAGTCTTACATCGCGCAGAGATGGTGC 681
Qy 685 TCGCGGAGATGATCAAGATTTGTGTGATACGCGGACCAAGAGATGTTCTCAAGGCC 744
Db 682 ACCCAAGACGCTAGACATGCTGTGAAGTGGGTGAACAAACGCCCCCTTGCTGAAAGTG 741
Qy 745 GCGCGCGGTTTTTGCATGATTTTTCGAGCCCGACGCGCGCGCTCGGAGCGGCTCGCG 804
Db 742 GCGCGCGGCGAGTTCCATGATTTTTCGCGCGGACGAGCGACACTGCTCTTACCTGCT 801
Qy 805 GAGACCGAAGAGGAGCTGCTGTGCTGCGGATATCGACCTCGGCAATGATCGGTTGGCCAG 864
Db 802 CACGATGCCAGGGGCTTATGATCTGCGGATCTGAATATGAGAGATTTGCTTCCGCAA 861
Qy 865 GCGCGCGCGCATCCGCGCGGCGCACTATTTCAAGCGCCGACGTAAGCGCGCTGCTGGA 923
Db 862 GCGATCAATGACCCCGTAGGCACTATTTCAAAACCGGAGCGACCCGCTGCTGCTGGA 920
```

```
RESULT 2
US-10-919-182-3
; Sequence 3, Application US/10919182
; Publication No. US2006003532A1
; GENERAL INFORMATION:
; APPLICANT: E.I. duPont de Nemours and Company, Inc.
; APPLICANT: Di Cosimo, Robert
; APPLICANT: Payne, Mark
; APPLICANT: O'Keefe, Daniel
; TITLE OF INVENTION: PRODUCTION OF 3-HYDROXYCARBOXYLIC ACID USING NITRILASE MUTANTS
; FILE REFERENCE: CL2584 US NA
; CURRENT APPLICATION NUMBER: US/10/919,182
; CURRENT FILING DATE: 2004-08-16
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 3
; LENGTH: 1110
; TYPE: DNA
; ORGANISM: Acidovorax facilis 72M
; FEATURE:
; NAME/KEY: CDS
```

LOCATION: (1)..(1110)  
US-10-919-182-3

Query Match 27.3%; Score 283.8; DB 9; Length 1110;  
Best Local Similarity 57.0%; Pred. No. 6.1e-62;  
Matches 519; Conservative 0; Mismatches 392; Indels 0; Gaps 0;

```
Qy 19 AAGTATTCGGCGCGCGGCGGTGACAGGCGCGCGGTGTTCTCGATTCGACCGCACAGTC 78
Db 19 AAGTCTCTCGCGCAACCGTTCAAGGACGCGCGGTATGCTTCACCGACGACGCAAGATC 78
Qy 79 GAGAAAGCATTCGAGCTGATCGAGCAGCGCGCAAGAGAGACGTGGCGCTGATCGCATTC 138
Db 79 GACAAAGTATTCGACATCATCGAAGAGCTGCCAAAGGCGCGAGTGTGATCGCTTTC 138
Qy 139 CCAAGACTTGGATTCGCGCTATCCCTTTGGATATGAGCTGGGCGCGCGCTTGGGCG 198
Db 139 CCGAAGTATTCATTCGCGGCTACCCCTATTTGGGCGGTGGCTGGCGACGTGAAGTACAG 198
Qy 199 ATGCGCTTGTCCAGCGCTATTTTGAAGATTTGCTGTGCGGCGACGACAGTGGCAG 258
Db 199 CTAGACTTATCTTACGCTATTCAGGAATTTGTTGAAGCTAGTGACGACCGTATGCGT 258
Qy 259 GCGCTGCGGATGCGGCGCGCGCGCGCACGCGCATGCTGCGCGGCTATAGCGAGCGC 318
Db 259 CGCTCAGGTGCGCGCGCGCGCGCAACAAATGCACTGTGATGGGCTATTCGAGCGG 318
Qy 319 GCGGCGGCGAGCTCTATATATGGGCGAGCGATTTTGGCGCGCGAGTGGCATTCATGCC 378
Db 319 GAAGCGGATTCGGCTATCTAGGCGAGGTTCATCGAAGAGGTGGCGAGATCGTTGCC 378
Qy 379 GCGCGCGCAAGCTCAAGCTACCCCATGGCGAGCGCACCGTGTGGCGAGGAGCGCG 438
Db 379 AATCGCGGAGGTGAAGCTCACACACGTTGAGCGTACATCTTACGCGAAGCGAAC 438
Qy 439 AGCATTCGCGGTGACGATACCGCATCGGAGCGCTTGGCGCGCTCTGTTGCTGGAG 498
Db 439 ACCGATTCCTACGACGACGATTCGCGGTTGCGAGCGGTGGTGAATTGAATGCTGGGAA 498
Qy 499 CACATCAAGCAATGTTCGAATAATACGATACGCGCGCGAGAGAACAGTCCACGTCGG 558
Db 499 CATTTCAACCGGTGCAAGTTGATGATGTACAGCTTGGTGAAGAGGTTCACGTTGCA 558
Qy 559 TCGTGGCGGAGCTTACGCTTATCGCGGACATGAGCTTATGCGCTCGGACCGGAGTCAAT 618
Db 559 TCGTGGCGGCGATGCTTCCCTTTCAGCGGATGTTTCCAACTGAGCATTCGAAGCCAA 618
Qy 619 ACCGCGCAAGCGATTTACGCGGTGAGGCGGCTGCTACGTGTGCGCGG 678
Db 619 GCGACGATCACCGGCTGTAACCAATCGAAGGCCAAACCTTGTGCTTGTGCTGACGCGAG 678
Qy 679 ACCGTTTGGCGGAGATGATCAAGGATGTGTGATACCGCGGACCAAGAGATGTTCTTC 738
Db 679 GTGATGAGCTTAGCGCGATCGAAGACGTTTCTCTTACGAGAGAACAGGCGCACGTGG 738
Qy 739 AAGGCGCGCGGCTTTTTCATGATTTTTCGAGCGCGACGCGCGCGCTTGGCGGCGG 798
Db 739 CGCAAGAGATGTGCTGGCGCGCATTTTACGCGCGCGGATGAGAGAGGCTTCCGAAGCT 798
Qy 799 CTCCGGAAGACGGAAGAGGAGCTGCTGTGCGCGATATGACCTTGGCATGATTCGGTTG 858
Db 799 CTGGCGGAAGATGTGAGGGGATCTTGTACGCAAGATGATCTGTAGAGAGATTCGCTG 858
Qy 859 GCAAGCGCGCGCGATCGCGCGGCGCATTTTACAGCGCGCGGAGTGAACGCGGCGCTG 918
Db 859 GCGAAGCTGAGGCGGATCGGTGGGCACTATTTGCGGCTGACGTGCTGTGGTCCAG 918
Qy 919 CTGATTCGAGCG 929
Db 919 TTCGACCGCGG 929
```

RESULT 3



Db 319 GAAGCCGAGTGGCGCTATCTGAGCCAGGTGTTTCATCGACGAGCGGCGAGATCGTTGCC 378  
Qy 379 GCGGCGCCGAGGCTCAACCTTAACCCATCGGAGCGCACCGTGTTCGCGAGGAGACGCGC 438  
Db 379 AATGGCGCAAGCTGAAACCCACACACGTTGAGCGATCTACGCGGAAAGGACGAGA 438  
Qy 439 AGCCATCTCGGCGTGCAGATATACGCGCATCGGCGCGCTCGGCGCTGTGTGCTGGGAG 498  
Db 439 ACCGATTTCTCAGCAGCATCTTGCCTTCGAGCGCGCTGGGAGTTGAACCTGCTGGAA 498  
Qy 499 CACATCCAGCCATTGTGAAATACGCCATGTACGCGCGGAGAGAACAGTCCAGTCCGCG 558  
Db 499 CATTTCCAAACCGCTCAGCAAGTTATGATGTACAGCTCGGTGAGCAGGTCCAGTTGCA 558  
Qy 559 TCGGCGCGGAGCTTCAAGCTCTATCGCGGCGATGCGCTTACGCGTCCGAGCCGAGTCAAT 618  
Db 559 TCGGCGCGGAGTGTCCCTCTTCAGCGGAGTGTTCGAACTGAGATCGAAGCCGAC 618  
Qy 619 ACCGCGCGAAGCCAGATCTACGCGGATCGAGCGCGGCTGTACGTTGCGTGGCGCG 678  
Db 619 GCGACGGTCTGCGCTGCTGTAAGCAATCGAAGGCCAAACCTTTGTGCTTGTCCAGCAG 678  
Qy 679 ACCGTTTCCGCGAGATGATCAAGTATGTGTGATACGCCCGAACAGAGATTTCTC 738  
Db 679 GTGATCGGACCTAGCGCGATGAAACGTTCTGCTCAACGACGAGACGCGCACTGTG 738  
Qy 739 AAGCGCGCGCGGCTTTTTCATGATTTTTCGCGCGCGCGCGCGCGCGCTGGCGAGCGG 798  
Db 739 CCGGAAAGATGTGCTGGCGCGCATTTACGCGCGCGATGAGAGCGAGCTTGGCAAGCTT 798  
Qy 799 CTCCCGAGAACGGAAGGAGCTGCTGCTGCGCGATATCGACTCGGCGATGATCGCGTTG 858  
Db 799 CTGCGGGAAGATGCTGAGGGGATCTTGTACGCAAGATCGATCTGAGCAGATTTCTGCTG 858  
Qy 859 GCCAAGCGGCGCGGATCGCGCGCGCGCACTATTCAGCGCGCGCGCGCGCGCGCTGCTG 918  
Db 859 GCCAAGCTGAGCGCGATCGCGTGGGCACTATTCGCGCGCTGAGCTGTGCTGCTCAG 918  
Qy 919 CTGGATCGACG 929  
Db 919 TTCGACCCGCG 929

RESULT 5  
US-10-919-182-5  
; Sequence 5, Application US/10919182  
; Publication No. US20060035352A1  
; GENERAL INFORMATION:  
; APPLICANT: E.I. duPont de Nemours and Company, Inc.  
; APPLICANT: Di Cosimo, Robert  
; APPLICANT: Payne, Mark  
; TITLE OF INVENTION: PRODUCTION OF 3-HYDROXYCARBOXYLIC ACID USING NITRILASE MUTANTS  
; FILE REFERENCE: CI2584 US NA  
; CURRENT APPLICATION NUMBER: US/10/919,182  
; NUMBER OF SEQ ID NOS: 18  
; SOFTWARE: Patentin version 3.2  
; SEQ ID NO 5  
; LENGTH: 1110  
; TYPE: DNA  
; ORGANISM: Artificial sequence  
; FEATURE:  
; OTHER INFORMATION: Mutant nitrilase B2 and H9  
; NAME/KEY: CDS  
; LOCATION: (1)..(1110)  
US-10-919-182-5

Query Match 27.1%; Score 282.2; DB 9; Length 1110;  
Best Local Similarity 56.9%; Pred. No. 1.5e-61;  
Matches 518; Conservative 0; Mismatches 393; Indels 0; Gaps 0;

Qy 19 AAGTATGCGGCGCGGCTGCAAGCGCGCGCGGCTTCTCGATCTCGACCGCACAGTC 78  
Db 19 AAGTCTCTCGCGCAACCGTTTCAGGAGAGCCCGATGTGCTCGACGACGACAGCATC 78  
Qy 79 GAGAAAGCATCGGCGCTGATCGAGCGGCGCAAGAGGAGCGTGGCGCTGATCGCATTC 138  
Db 79 GACAACTGATCGGCTGATCGAAGAACTGCCAAAGGCGCGAGTGTGATGCTTTC 138  
Qy 139 CCAAGACTTGATTCGCGCTATCCCTTTTGGATATGCTGCGCGCGCGCGCTGGGCG 198  
Db 139 CCGAAGATATTCATTCGCGGCTACCCCTATTTGGCGCTGCTGCGCGACGTGAAGTA 198  
Qy 199 ATGCGCTCTCTCAAGGCTATTTTCGAATTTGCTGTGCGCGCGAGCAAGCTGCGAG 258  
Db 199 CTAACTTATCTTACAGCTATCAAGAAATTTGTGAGCTAGATGACGACCGTATGCGT 258  
Qy 259 GCCCTGCGGAGTCCGCGCGCGCGCGCGCATGCTGCGCGCGCTATATGAGAGCGC 318  
Db 259 CGCTCCAGCTGCGCGCGCGCGCGCGCAAMAATGCACTGCTCATGCGCTATTCGAGCGG 318  
Qy 319 GCGGCGGAGCGCTCTATATGCGCGAGCGATTTGCGCGCGAGTCCGATCTGATCGCT 378  
Db 319 GAAGCGGATCGGCTATCTGAGCGAGGCTGATTCAGAGCGGCGAGATCGTTGCC 378  
Qy 379 GCGCGCGCAAGCTCAAGCTTACCATGCGAGCGGACGCGTGTTCGCGGAGGAGAGCGG 438  
Db 379 AATCGCGCAAGCTTGAAGCCACACAGCTTGAAGCTATCATCGAGGCGAAGGAA 438  
Qy 439 AGCCATCTCGCGGTGACAGATACCGCATCGGCGCGCTGCGCGCTGTGCTGCGAG 498  
Db 439 ACCGATTTCTTCCAGCAGCATCTTCCGCTTCGAGCGCGTGGATTTGAATCTGCTGGAA 498  
Qy 499 CACATCAGCGATGTCGAAATACGCCATGTACGCCCGCGAGAGACAGTCCAGCTGCGG 558  
Db 499 CATTTCAACCGCTCAGCAAGTTATGATGTACAGCTGTGAGGAGGTCCAGTTGCA 558  
Qy 559 TCGTGGCGAGCTTCAAGCTTATCGCGGATGCGCTATGCGCTCGGACCGAGGTCAAT 618  
Db 559 TCGTGGCGGCGATGTCCTCTTTCAGCGCGGATGTTTTCATCTGAGCATTCGAAGCAAC 618  
Qy 619 ACCGCGCGAAGCCAGATCTACCGCGTTCAGAGCGCGCTGTAAGTGTGCTGCGCG 678  
Db 619 GCGAGCTGCGCGCTCGTTCGATCGAATCGAAGCCAAACCTTTGTGCTTGTGCTGACG 678  
Qy 679 ACCGTTTCCGCGAGATGATCAAGGTAATGTGATGATACGCCCGCAAGAGATGTTCTC 738  
Db 679 GTGATCGACCTAGCGCGATCGAAGCTTCTGCTCAAGACGACGCGGACGCTGTTG 738  
Qy 739 AAGCGCGCGCGGTTTTCATGATTTTTCGAGCGCGCGCGCGCGCGCTGCGCGAGCGG 798  
Db 739 CCGCAAGATGTGCTGCGCGCGCATTTTACCGCGCGGATGGAAGCGAGCTTCGAAGCTT 798  
Qy 799 CTCCCGGAACGGAAGGAGCATGCTGTGTCGCGGATTCGAGCTCGGCTATGCTGCTG 858  
Db 799 CTGCGGGAAGATGCTGAGGGGATCTTGTACGAGAGATGATCTGAGCAGATTTCTGCTG 858  
Qy 859 GCCAAGCGCGCGCGATCGCGCGCGCGCACTATTCAGCGCGCGCGAGATACGCGGCTGCTG 918  
Db 859 GCCAAGCTGAGCGCGATCGCGTGGGCACTATTTGCGCGCGCTGAGGTGCTGTGCTGCTCAG 918  
Qy 919 CTGGATCGACG 929  
Db 919 TTCGACCCGCG 929

RESULT 6  
US-10-919-182-13  
; Sequence 13, Application US/10919182  
; Publication No. US20060035352A1  
; GENERAL INFORMATION:  
; APPLICANT: E.I. duPont de Nemours and Company, Inc.  
; APPLICANT: Di Cosimo, Robert  
; APPLICANT: Payne, Mark

```

; APPLICANT: O'Keefe, Daniel
; TITLE OF INVENTION: PRODUCTION OF 3-HYDROXYCARBOXYLIC ACID USING NITRILASE MUTANTS
; FILE REFERENCE: CI2584 US NA
; CURRENT APPLICATION NUMBER: US/10/919,182
; CURRENT FILING DATE: 2004-08-16
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 13
; LENGTH: 1110
; TYPE: DNA
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Mutant nitrilase with Phe168 to Lys change
; NAME/KEY: CDS
; LOCATION: (1)..(1110)
; US-10-919-182-13

Query Match      27.1%; Score 282.2; DB 9; Length 1110;
Best Local Similarity 56.9%; Pred. No. 1.5e-61;
Matches 518; Conservative 0; Mismatches 393; Indels 0; Gaps 0;

Qy 19 AAGTATCGCGCGCGCGGTCAGAGCGCGCGGTCCTCGATCTCGACCGCAGATC 78
Db 19 AAGTCTCTCGCGCAACGGTTCAGAGCAGAGCGGTCGTCGACGACGACGACATC 78
Qy 79 GAGAAAGCATCGGCTGATGAGCAGAGCGGCGCAAGACGATGCGCTGATGCAATTC 138
Db 79 GACAAATGATCGGATCATGAGAAAGCTCCAAAAGGCGCGAGATGATGCTTTC 138
Qy 139 CCGAGACTTGATTCCTCCGCTATCCCTTTGATATGCTGGGCGCGGCTTGGGGC 198
Db 139 CCGGAATATATCATCTCGGCTACCCCTATGCGCGTGGCTCGCGCATGAATACAGC 198
Qy 199 ATGCGCTTCGTCAGCGCTATATTCAGAAATTCGCTGTCGCGCGCAGCAAGCAGTGGCAG 258
Db 199 CTAAAGCTTTACTTACCGCTATCAAGAAATTCGTTGAGTAGTGAAGACCGTATGCT 258
Qy 259 GCCCTGCGGATGCGCGCGCGCGCGCAGCATGCTGTCGCGCTATAGCGAGCGC 318
Db 259 CGCCTCGAGCTGCGCGCGCGCGCGCAAAATCCACATCGCTATGCGGCTATTCGGAACGG 318
Qy 319 GCGCGCGCGAGCTCTATATGAGCGCGCATGCTTTCGCGCGCGCGCGCATGATGCGC 378
Db 319 GAAGCCGATGCGCGCTATGAGCGCGATGCTTTCATGCAAGCGCTGCGAGATGTTGCC 378
Qy 379 GCGCGCGCGAGCTCAAGCGCTTACCATGCGGAGCGCAACCGTTCGCGGAGGAGAGCGC 438
Db 379 AATGCGGCGAGCTGAAGCGCGCACAGCTTGAAGCTGATGATCTACGCGGAGGAGCGA 438
Qy 439 AGCCATCTCGGATGAGCATACCGCATCGGCGCGCTCGCGCGCTCTGTTGCTGGAG 498
Db 439 ACCGATTTCTCAGCAGCATCTTGCCTTGCAGCGCGCTGAGATTGAATGCTGGGAA 498
Qy 499 CACATCAAGCATGTGGAATTAAGCATGTAACCGCGCGCAAGACGATGCAAGTGGCG 558
Db 499 CATTAACAACCGCTCAGCAAGTTGATGATGATGAGCTGCGTGAAGAGTCAAGTTGCA 558
Qy 559 TCGTGGCGGAGCTTCAAGCTTATCGCGGCGATGCGCTATGCGCTCGGACCGAGATGAAT 618
Db 559 TCGTGGCGGAGATGATCCCTCTTTCAGCGCGGATGTTTTCACATGAGATGAAAGCCAC 618
Qy 619 ACCGCGCGCAAGCATCTACGCGGTGAGGCGGCTGCTACGCTGCGCTGCTGCGCG 678
Db 619 GCGAGGTCACCGCTCGTACGCAATGGAAGCGCAAACTTTGCTTGGCTTGCACGAG 678
Qy 679 ACCGCTTCGCGGAGATGATGAAGTATGATGATGATGATGATGATGATGATGATGATG 738
Db 679 GTGATGCACTTACGCGGATGAAACGTTCTGCTCAAGCAAGCAAGCGGCACTGTTG 738
Qy 739 AAGCGCGCGCGGCTTTTTCATGATTTTTCGCGCGCGCAAGCGCGCGCTGCGCGAGCG 798
Db 739 CCGGAAGATGATGCTGCGCGCGCATTTACGCGCGCGATGGAACGAGCTTGGCAAGCTT 798
```

```

Qy 799 CTCGCCGAGACCGAAGAGGAGCTGCTGTCGCGGATATGACCTCGGCATGATCGCGTTG 858
Db 799 CTCGCCGAGATGAGGAGATCTTGTACGAGAGATGATGAGGAGATTCCTG 858
Qy 859 GCCAAGCGCGCGCGCGATCCGCGCGCGCACTATTCAGCGCGCGCAAGTAAAGCGCTGCG 918
Db 859 GCGAAGCTGAGAGCGGATCCGCGCGCGCATTTTCGCGCGCGCTGACGTCGTGCTCAG 918
Qy 919 CTGATGACG 929
Db 919 TTCGACCGCG 929

RESULT 7
US-10-919-182-15
; Sequence 15, Application US/10919182
; Publication No. US20060035352A1
; GENERAL INFORMATION:
; APPLICANT: E.I. duPont de Nemours and Company, Inc.
; APPLICANT: DI Cosimo, Robert
; APPLICANT: Payne, Mark
; APPLICANT: O'Keefe, Daniel
; TITLE OF INVENTION: PRODUCTION OF 3-HYDROXYCARBOXYLIC ACID USING NITRILASE MUTANTS
; FILE REFERENCE: CI2584 US NA
; CURRENT APPLICATION NUMBER: US/10/919,182
; CURRENT FILING DATE: 2004-08-16
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 15
; LENGTH: 1110
; TYPE: DNA
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Mutant nitrilase with Phe168 to Val change
; NAME/KEY: CDS
; LOCATION: (1)..(1110)
; US-10-919-182-15

Query Match      27.1%; Score 282.2; DB 9; Length 1110;
Best Local Similarity 56.9%; Pred. No. 1.5e-61;
Matches 518; Conservative 0; Mismatches 393; Indels 0; Gaps 0;

Qy 19 AAGTATCGCGCGCGGTCAGAGCGCGCGGTCCTCGATCTCGACCGCAGATC 78
Db 19 AAGTCTCTCGCGCAACGGTTCAGAGCAGAGCGGTCGTCGACGACGACGATC 78
Qy 79 GAGAAAGCATCGGCTGATGAGCAGAGCGGCGCAAGACGATGCGCTGATGCAATTC 138
Db 79 GACAAATGATCGGATCATGAGAAAGCTCCAAAAGGCGCGAGATGATGCTTTC 138
Qy 139 CCGAGACTTGATTCCTCCGCTATCCCTTTGATATGCTGGGCGCGCGCTTGGGGC 198
Db 139 CCGGAATATATCATCTCGGCTACCCCTATGCGCGTGGCTCGCGCATGAATACAGC 198
Qy 199 ATGCGCTTCGTCAGCGCTATATTCAGAAATTCGCTGTCGCGCGCGCAAGCAGTGGCAG 258
Db 199 CTAAAGCTTTACTTACGCTATCAAGAAATTCGTTGAGTAGTGAAGACCGTATGCT 258
Qy 259 GCCCTGCGGATGCGCGCGCGCGCGCGCATGCAATGCTGTCGCGCTTATGAGAGCGC 318
Db 259 CGCCTCGAGCTGCGCGCGCGCGCGCAAAATCGCATGCTGATGAGCTATTCGAGCGG 318
Qy 319 GCGCGCGCGAGCTCTATATGAGCGCGAGCGATCTTTCGCGCGCGCGATGATGCGC 378
Db 319 GAAGCGGATGCGGCTATGAGCGCGAGCTTTCATGACAGAGGTCGCGGAGATGCTTGC 378
Qy 379 GCGCGCGCAAGCTCAAGCTTACCATGCGAGCGCACGCTGTTGCGGAGGAGAGCGC 438
Db 379 AATGCGGCGAAGCTGAAGCGCGCACAGCTTGAAGCTGATGATGATGATGATGATGATG 438
Qy 439 AGCCATCTCGGATGAGCATACCGCATGCGGCGCGCTCGCGCGCTCTGTTGCTGGAG 498
```

```
Db 439 ACCGATTTCTCAGCAGCAGCTTCCGCTCCGAGCCGCTGGATTTGAACCTGCTGGGAA 498
Qy 499 CACATCCAGCATTTGTGAAATACGCCATGTACCGCCGCGAGAACAGTCCAGCTGGCG 558
Db 499 CATTTTCAACCGCTCAGCAATTTATGATGTAACAGCTTCGGTGAAGAGTTCAGATTGCA 558
Qy 559 TCGTGGCCGAGCTTCAAGCTCTATCCGCGGCAATGAGCTTATGCGTCCGAAACCGAGAGTCAAT 618
Db 559 TCGTGGCCGCGAGTGTCCCTCTTCAAGCCGATGTTTTTCCAACTGAGCATGAAAGCAAC 618
Qy 619 ACCGCCGCAAGCCAGATCTACAGCGGTGAGGGCGGAGCTGCTACAGTCTGAGCTGCTGGCGG 678
Db 619 GCGAGGTCACCCCTGCTGACGCAATGCAAGGCGCAAACTTTGTGCTTTGCTGCAAGCAG 678
Qy 679 ACCGTTTCGCGGAGATGATCAAGATATTGTGTGATACGCCCGAACAGAGATTTCTTC 738
Db 679 GTGATCGGACCTTACCGGAGTGAAGAGTTCTGCTCAACGACGAAACAGCCGCACTGTG 738
Qy 739 AAGGCCGCGCGGTTTTCCTATGATTTTCCGCGCCGAGCGCGCCCTGGCCGAGCCG 798
Db 739 CCGGAAAGATGTGCTGAGCGCGCATTTTACGCCGAGTGAAGCGAGCTTGGCGAAGCCT 798
Qy 799 CTCGCGGAGACCGAAGAGGAGCTGTGATGATTTTCCGCGATATCGACTCGGATGATCGCGTTG 858
Db 799 CTGGCGGAAAGATGTGATGAGGAGTCTTGTACGCAAGATCGATCTGAGCAGATTTCTGCTG 858
Qy 859 GCCAAGCGCGCGCGCATTCGCGCGGCGCACTATTACGCGCCGAGCGTAAACGCGCTGCTG 918
Db 859 GCGAAGGCTGGAAGCGGATCCGGTGGGCACTATTCGCGGCTGACGTGCTGCTGCTCAG 918
Qy 919 CTGGATCGACG 929
Db 919 TTGACCCGCG 929
```

```
RESULT 8
US-10-919-182-17
; Sequence 17, Application US/10919182
; Publication No. US200600352A1
; GENERAL INFORMATION:
; APPLICANT: E.I. duPont de Nemours and Company, Inc.
; APPLICANT: DI Cosimo, Robert
; APPLICANT: Payne, Mark
; APPLICANT: O'Keefe, Daniel
; TITLE OF INVENTION: PRODUCTION OF 3-HYDROXYCARBOXYLIC ACID USING NITRILASE MUTANTS
; FILE REFERENCE: CI2584 US NA
; CURRENT APPLICATION NUMBER: US/10/919,182
; CURRENT FILING DATE: 2004-08-16
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 17
; LENGTH: 1110
; TYPE: DNA
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Mutant nitrilase with Phe168 to Leu change
; NAME/KEY: CDS
; LOCATION: (1)..(1110)
US-10-919-182-17
```

```
Query Match 27.1%; Score 282.2; DB 9; Length 1110;
Best Local Similarity 56.9%; Pred. No. 1.5e-61;
Matches 518; Conservative 0; Mismatches 393; Indels 0; Gaps 0;
```

```
Qy 19 AAGTATCGCGCGCGCGGCGGCGGCGCGCGGCTTCTCGATCTCGACCGCAAGTC 78
Db 19 AAGTTCCTCGCGCAACCGTTACGAGAGCCGATATGCTTCACGCAAGCAAGATC 78
Qy 79 GAGAAAGCATCGCGCTGATCGAGCAGCGCCCAAGAGAGACGTCGCTGATCGCAATG 138
Db 79 GACAAAGTCATCGGATCATCGAAGAGCTGCCCAAAAGGCGCGAGTCTGATCGCTTTC 138
```

```
Qy 139 CCAAGACTTGATATCCCGGCTATCCCTTTGGATATAGGTGAGCGCGGCTTGGGGC 198
Db 139 CCGAAGATATTCATTCGGGCTACCCCTATTTGGCGTGGCTCGGACGTAAGTACAGC 198
Qy 199 ATGCGCTTTCGTCAGCGCTATTTTGAAGATTTGCTGTGCGGCAAGCAAGCTGGCAG 258
Db 199 CTAAAGCTTACTTACACCTATTCAGAAATTCGTTGAGCTTAAGTGAACGACCTGATCGT 258
Qy 259 GCCCTGCGGATGCGCGCGCGCGCGCACGCGATGATGTGCGGCGCGGCTTTTAAAGACGC 318
Db 259 CGCTTCAGCTGCGCGCGCGCGCGCACAAATGCACTGTCATGAGCTTATTTGGAACGG 318
Qy 319 GCGGCGGAGCCTCTATATGAGCGAGCGATCTTGGCCCGATGCGATGCGATCGCC 378
Db 319 GAAGCGGATTCGCGCTATCTGAGCGAGGTGTTATGACAGAGGTGCGAGATGTTGCC 378
Qy 379 GCGGCGGCAAGCTCAAGCTTACCCATGCGGAGCGCACCGTGTTCGCGAGGAGACGCC 438
Db 379 AATCGGCGCAAGCTGAAGCCCAACACGTTGAGCGTACGATCTACGCGCAAGGCAACGA 438
Qy 439 AGCCATCTCGCGGTGACGATACCGCCATCGGCGCCTCGCGCGCTCTGTGCTGGAG 498
Db 439 ACCGATTTCTCAGCAGCAGCTTCCGCTTGGACGCGTCCGCTGATTTGAACCTGCGGAA 498
Qy 499 CACATCCAGCATTTGTGAAATACGCCATGACCGCGCGAGCAAGACAGTTCAGCTGCG 558
Db 499 CATCTACACCGCTCAGCAAGTTCATGATGTAACGCTCGTGAAGCAGTTCAGTTCGA 558
Qy 559 TCGTGGCCGAGCTTACGCGCTCTATCGCGGCAATGCGCTGCGGACCGGAGGTCAAT 618
Db 559 TCGTGGCCGCGAGTGTCCCTCTTCAAGCCGAGATGTTTTCCAATGAGCATGCAAGCCAC 618
Qy 619 ACCGCCGAAAGCCAGATCTACGCGGTGAGGCGGCGCTCTACGCTGCGCTGCGCG 678
Db 619 GCGAGGTCACCCGCTGCTGACGCAATCGAAGGCGCAAACTTTGCTTGTCTCGACGAG 678
Qy 679 ACCGTTTCGCGGAGATGATCAAGGATTTGTGTGATATCGCCCGAACAGAGATGTTCTTC 738
Db 679 GTGATGGAATCTTACCGCGATCGAAGCTTGTGCTTCAACGACAGCGCGCATCTGTG 738
Qy 739 AAGCGCGCGCGCGGCTTTTGTGCATGATTTTTCGCGCCGACGCGCGCCCTGCGCGAGCCG 798
Db 739 CCGCAAGATGTGCTGAGGCGCATTTTACGCGCCGAGTGAAGCGAGCTTGGAGGCTT 798
Qy 799 CTCGCGGAGACCGAAGAGGAGCTGTGTGCGCGCGCGCATATGCACTCGCATGATCGCGTTG 858
Db 799 CTGGCGGAAAGATGCTAAGGGGATCTTGTACGCAAGATCGATCTGAGCAGATTTCTGCTG 858
Qy 859 GCCAAGCGCGCGCGCATTCGCGCGGCGCACTATTTCACGCGCGAGCGTAAACGCGGCTGCTG 918
Db 859 GCGAAGGCTGGAAGCCGATCCGGTGGGCACTATTTCGCGGCTGACGTGCTGCTGCTCAG 918
Qy 919 CTGGATCGACG 929
Db 919 TTGACCCGCG 929
```

```
RESULT 9
US-10-496-351-4
; Sequence 4, Application US/10496351
; Publication No. US20060084141A1
; GENERAL INFORMATION:
; APPLICANT: Floss, Heinz
; APPLICANT: Yu, Tin-Wei
; APPLICANT: Leister, Eckard
; TITLE OF INVENTION: Biosynthetic Gene Cluster for the Maytansinoid Antitumor Agent
; FILE REFERENCE: UMASH-06712
; CURRENT APPLICATION NUMBER: US/10/496,351
; CURRENT FILING DATE: 2004-05-19
; NUMBER OF SEQ ID NOS: 57
; SOFTWARE: PatentIn version 3.0
```



```
; SEQ ID NO 4
; LENGTH: 9975
; TYPE: DNA
; ORGANISM: Actinomyces pretiosum
US-10-496-351-4
```

Query Match 4.9%; Score 50.6; DB 7; Length 9975;

Best Local Similarity 43.7%; Pred. No. 0.0025; Mismatches 409; Indels 9; Gaps 2;

Matches 325; Conservative 0; Mismatches 409; Indels 9; Gaps 2;

```
Qy 186 GCCGCTTGGGCGATGCGCTTCTGTCAGCGCTATTTCAGAAATTCGCTGTCGCGCGAG 245
Db 3588 GCCGCTGCTGGCCAGCGATGTCCTCCGCGGCTCCGCGAGGACGCGGTGTCCACCT 3647
Qy 246 CAAGCATGCGAGCGCTTGGCGGATGCGCCCGCCAGCGCATGTCGTGCGCG 305
Db 3648 CGCGGTGTGCGAGGGGTTCCTGCGGATGTCGCGCGACCGCATCTGCGTCTGTGT 3707
Qy 306 CTATAGCGAGCGCGCGCGGCGAGCTCTATATGCGCGAGGAGATCTTCCGCGCGGATG 365
Db 3708 CACCAAGGGGCGCGAGGCGCGCGAGCGCGCGAGCGCGCGAGCGCGCGCGCGCT 3767
Qy 366 CGATCTGATGCGCGCGCGCGCGAGCTCAAGCTTACCGCATGCGAGCGCGCGTTCG 425
Db 3768 CGTCCGCTCCGCGCGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTC 3827
Qy 426 CGAGGAGACGCGCGAGCGCATCTGCGCGTGCAGATACCGCATCGGCGCGCTCGCGCT 485
Db 3828 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTC 3887
Qy 486 CTGTTGCGGAGGACATCCAGCGCATTCGAAATAGC---CGATGACGCGCGCGAG 542
Db 3888 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 3947
Qy 543 ACAGGTCCAGCTGCGCTGTCGCGCGAGCTTACGCTTATCGCGCGCATGCGCTTACGCT 602
Db 3948 CTTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 4007
Qy 603 CGGACCGGAGGTCAATACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 662
Db 4008 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 4062
Qy 663 GCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 722
Db 4063 -CGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 4121
Qy 723 CAAGGATGTTCTCTCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 782
Db 4122 CCGCGGTGTGTCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTGA 4181
Qy 783 CGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 842
Db 4182 CGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTGA 4241
Qy 843 CGGATGATGCGCGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 902
Db 4242 CGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 4301
Qy 903 CGTAAAGCGCGCTGCTGATC 925
Db 4302 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 4324
```

RESULT 10  
US-10-496-351-56/c

; Sequence 56, Application US/10496351

; Publication No. US20060084141A1

; GENERAL INFORMATION:

; APPLICANT: Floss, Heinz

; APPLICANT: Yu, Tin-Wei

; APPLICANT: Leisner, Eckard

; TITLE OF INVENTION: Bioanalytic Gene Cluster for the Maytansinoid Antitumor Agent

; TITLE OF INVENTION: Ansamitocin

```
; FILE REFERENCE: UMSH-06712
; CURRENT APPLICATION NUMBER: US/10/496,351
; CURRENT FILING DATE: 2004-05-19
; NUMBER OF SEQ ID NOS: 57
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 56
; LENGTH: 82746
; TYPE: DNA
; ORGANISM: Actinomyces pretiosum
US-10-496-351-56
```

Query Match 4.9%; Score 50.6; DB 7; Length 82746;

Best Local Similarity 43.7%; Pred. No. 0.0023; Mismatches 409; Indels 9; Gaps 2;

Matches 325; Conservative 0; Mismatches 409; Indels 9; Gaps 2;

```
Qy 186 GCCGCTTGGGCGATGCGCTTCTGTCAGCGCTATTTCAGAAATTCGCTGTCGCGCGAG 245
Db 63613 GCCGCTGCTGGCCAGCGATGTCCTCCGCGGCTCCGCGAGGACGCGGTGTCCACCT 63554
Qy 246 CAAGCATGCGAGCGCTTGGCGGATGCGCCCGCCAGCGCATGTCGTGCGCG 305
Db 63553 CGCGGTGTGCGAGGGGTTCCTGCGGATGTCGCGCGACCGCATCTGCGTCTGTGT 63494
Qy 306 CTATAGCGAGCGCGCGCGCGCGAGCTCTATATGCGCGAGGAGATCTTCCGCGCGGATG 365
Db 63493 CACCAAGGGGCGCGAGGCGCGCGAGCGCGCGAGCGCGCGAGCGCGCGCGCGCT 63434
Qy 366 CGATCTGATGCGCGCGCGCGCGAGCTCAAGCTTACCGCATGCGAGCGCGTTCG 425
Db 63433 CGTCCGCTCCGCGCGCGCGAGCGCGCGCGCGCGCGCGCGCGCGCGCGCTC 63374
Qy 426 CGAGGAGACGCGCGAGCGCATCTGCGGTGCAGATACCGCATCGGCGCGCTCGCGCT 485
Db 63373 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 63314
Qy 486 CTGTTGCGGAGGACATCCAGCGCATTCGAAATAGC---CGATGACGCGCGCGAG 542
Db 63313 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 63254
Qy 543 ACAGGTCCAGCTGCGCTGTCGCGCGAGCTTACGCTTATCGCGCGCATGCGCTTACGCT 602
Db 63253 CTTGACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 63194
Qy 603 CGGACCGGAGGTCAATACCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 662
Db 63193 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 63139
Qy 663 GCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 722
Db 63138 -CGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 63080
Qy 723 CAAGGATGTTCTCTCAAGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 782
Db 63079 CCGCGGTGTGTCTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTGA 63020
Qy 783 CGCGCTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 842
Db 63019 CGCGGTGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCTGA 62960
Qy 843 CGGATGATGCGCGTTCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 902
Db 62959 CGGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 62900
Qy 903 CGTAAAGCGCGCTGCTGATC 925
Db 62899 CGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCGCT 62877
```

RESULT 11  
US-11-096-568A-20685

; Sequence 20685, Application US/11096568A

; Publication No. US20060048240A1

; GENERAL INFORMATION:





**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: April 27, 2006, 00:55:21 ; Search time 25.836 Seconds  
(without alignments)  
1288.550 Million cell updates/sec

Title: US-09-751-299-2

Perfect score: 1806

Sequence: 1 MSEPMKTYRGAAGVAAVFL.....EDKGDAPALRVAAESAAQAQ 346

Scoring table: BIOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Database : Listing first 45 summaries

1: p1r1:.\*  
2: p1r2:.\*  
3: p1r3:.\*  
4: p1r4:.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	869	48.1	356	2 A47181	nitrilase (EC 3.5.
2	832.5	46.1	366	2 A45070	nitrilase - Rhodoc
3	825.5	45.7	383	2 A43470	aliphatic nitrilas
4	616	34.1	349	1 A28658	nitrilase (EC 3.5.
5	592.5	32.8	354	2 JC4212	nitrilase (EC 3.5.
6	584.5	32.4	346	2 S77025	nitrilase (EC 3.5.
7	562	31.1	368	2 UQ1613	cyanide hydratase
8	468.5	25.9	346	2 T49148	nitrilase (EC 3.5.
9	467	25.7	346	2 T49147	nitrilase (EC 3.5.
10	464	25.7	339	2 S1969	nitrilase (EC 3.5.
11	463	25.6	339	2 T52259	nitrilase (EC 3.5.
12	461	25.5	346	2 S22398	nitrilase (EC 3.5.
13	456	25.2	339	2 T52262	nitrilase (EC 3.5.
14	420	23.3	362	2 T52266	nitrilase-1like pro
15	411	22.8	348	2 T03739	nitrilase (EC 3.5.
16	404.5	22.4	349	2 T03736	nitrilase (EC 3.5.
17	403.5	22.3	305	2 T27679	probable nitrilase
18	399.5	22.1	355	2 T52265	nitrilase (EC 3.5.
19	224.5	12.4	199	2 S50363	nitrilase homolog
20	199.5	11.0	579	2 AH3225	amidohydrolase [im
21	198	11.0	262	2 C75051	hydrolyase related
22	196	10.9	262	2 C71109	hydrolyase related
23	182	10.1	318	2 T48563	hypothetical prote
24	154	8.5	272	2 T41662	probable nitrilase
25	151	8.4	267	2 B72408	conserved hydrolyet
26	147.5	8.2	290	2 B81369	probable hydrolyet
27	144	8.0	259	2 E69863	conserved hydrolyet
28	143.5	7.9	297	2 F75263	probable hydrolyase
29	143	7.9	322	2 T38399	probable amidohydr

30	141.5	7.8	264	2 T36488	probable hydrolase
31	141.5	7.8	292	2 C87275	hydrolase, carbon-
32	139	7.7	302	2 A72510	hypothetical prote
33	137.5	7.6	292	2 E64558	conserved hypotnet
34	137.5	7.6	294	2 A71949	hypothetical prote
35	133.5	7.4	284	2 AD3275	beta-ureidopropion
36	132.5	7.3	271	2 H83195	conserved hypotnet
37	130	7.2	292	2 C83608	probable hydratase
38	130	7.2	440	2 T43198	nitrilase/Phit pro
39	129.5	7.2	293	2 G75274	nitrilase-related
40	129	7.1	576	2 G72277	NH(3)-dependent NA
41	128	7.1	294	2 AB0115	probable carbon-ni
42	126	7.0	257	2 C69264	conserved hypotnet
43	126	7.0	280	2 T28684	hypothetical prote
44	125	6.9	220	2 A84673	probable nitrilase
45	124.5	6.9	285	2 S58240	hypothetical prote

## ALIGNMENTS

```
RESULT 1
nitrilase (EC 3.5.5.1), arylacetone-specific - Alcaligenes faecalis
C:Species: Alcaligenes faecalis
C>Date: 21-Sep-1993 #sequence_revision 18-Nov-1994 #text_change 31-Dec-2004
C:Accession: A47181; S13860
R: Kobayashi, M.; Izui, H.; Nagasawa, T.; Yamada, H.
Proc. Natl. Acad. Sci. U.S.A. 90, 247-251, 1993
A:Title: Nitrilase in biosynthesis of the plant hormone indole-3-acetic acid from indole
A:Reference number: A47181; MUID:93126352; PMID:8419930
A:Contents: UM3
A:Accession: A47181
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-356 <KOB>
A:Cross-references: UNIPROT:P20960; UNIPARC:UPI000033514; GB:D13419; NID:G216202; PIDN:
A>Note: sequence extracted from NCBI backbone (NCBI:122081, NCBI:P122082)
R: Nagasawa, T.; Mauger, J.; Yamada, H.
Eur. J. Biochem. 194, 765-772, 1990
A:Title: A novel nitrilase, arylacetone nitrilase, of Alcaligenes faecalis UM3. Purification
A:Reference number: S13860; MUID:91099356; PMID:2269298
A:Accession: S13860
A:Status: preliminary
A:Molecule type: protein
A:Residues: 1-33 <NAG>
A:Cross-references: UNIPARC:UPI0000175E31
C:Superfamily: nitrilase (carbon-nitrogen hydrolase)
C:Keywords: hydrolase

Query Match 48.1%; Score 869; DB 2; Length 356;
Best Local Similarity 50.6%; Pred. No. 4.8e-66;
Matches 172; Conservative 50; Mismatches 108; Indels 10; Gaps 2;

QY 9 RGAAGAAATPFLDTRVKAIGLIDQAKODVLLAFPTTTPGVFWTLGAPAGMR 68
DB 8 RAAAVQAASPNYLDATGVDTTTELARQARDEGCDLVFGFTWLPQYFPHWALPAASL 67
QY 69 FVQRYFENSILVRGSKOMALDAPARRGMHVAGYSERAGSGLYMGQALFGPDGLTAAR 128
DB 68 YSARVYANSLSDSAERFORIAQAARTLGRIALGYSERSGSLYLQGLLDDGQMTMSR 127
QY 129 RKLKPTHAERTVFEGSDGSHLAVHDTAIGRLGALCWEHIQPLSKYAMVAADQVHVASW 188
DB 128 RKLKPTHAERTVFEGSDGSHLAVHDTAIGRLGALCWEHIQPLSKYAMVAADQVHVASW 187
QY 189 PPSSTLYRGMAAYALGPVNTAAASQIYAVGEGCYTLAASCATYSPPMKVLTVPKEMTLKA 248
DB 188 PPSSTLYRGMAAYALGPVNTAAASQIYAVGEGCYTLAASCATYSPPMKVLTVPKEMTLKA 247
QY 249 GGGFAMIFGPDGALAEPLPETEGLLVADIDLGIMLAKAAADPAGHYRPPVTRLLLD 308
DB 248 GGGSSMIFAPDGGTLAPYLPHDAEGLIADLNMBEIAFAKAIINDPVGHYSKPEATRLVLD 307
```

QY 309 ---RRPAQRVVTLDAAFEQNEDEKDAAPALRVVAESAAA 345  
Db 308 LGHREPTVTRHAKSVIOE-----EAFEPHVGSTAAFPVA 340

## RESULT 2

A45070  
nitrlase - Rhodococcus rhodochrous  
C/Species: Rhodococcus rhodochrous  
C/Date: 10-Jun-1993 #sequence\_revision 18-Nov-1994 #text\_change 31-Dec-2004  
C/Accession: A45070  
R/Kobayashi, M.; Kameda, H.; Yanaka, N.; Nagasawa, T.; Yamada, H.  
J. Biol. Chem. 267, 20746-20751, 1992  
A/Title: Nitrlase from Rhodococcus rhodochrous J1. Sequencing and overexpression of the  
A/Reference number: A45070; MUID:93015976; PMID:11400390  
A/Contents: J1  
A/Accession: A45070  
A/Status: preliminary  
A/Molecule type: DNA; protein  
A/Residues: 1-366 <KOB>  
A/Cross-references: UNIPROT:Q03217; UNIPARC:UPI000016FP36; GB:D11425; NID:9216933; PIDN:  
A/Note: sequence extracted from NCBI backbone (NCBIN:116039, NCBIT:116040)  
C/Superfamily: nitrlase (carbon-nitrogen hydrolase)

Query Match 46.1%; Score 832.5; DB 2; Length 366;  
Best Local Similarity 52.2%; Pred. No. 6.2e-63;  
Matches 166; Conservative 43; Mismatches 108; Indels 1; Gaps 1;

QY 1 MSEPMTRYKGAAYVFLDLDRVTEKAIGLEQAKODVRLIAPETWIPGYPMW 60  
Db 1 MVEYNTNFKYAAVQAPWFDAATVKTYSIIAEAAANGCELAFFEVLPFGYPIHW 60  
QY 61 GAPAMGM-RFVQRYFENSILVSGSKOMQALADAAARRHGMVAVAGYSEKAGSLYMGQIFG 119  
Db 61 DSIPLAGAKAVRNRHENSIMDSPHVQLDLDAADHNIIVVGGISERGGSLYMTQVYID 120  
QY 120 PDGDLIAARRKLTPTAERTVFGSGDSHLAVHDTAIGRLALCMWHIOPLSKYAMVA 179  
Db 121 ADGQLVARRRKLKPTHERSVYGGNGSDISVYMPARLALCMWHEFQTLTYKAMSM 180  
QY 180 DEQHVAVSWPSFSIYRGAAYALGPEVNTAASQIYAVEGCVYLASCATVSEPMIKVLVD 239  
Db 181 HEQHVAVSWPSFSIYRGAAYALGPEVNTAASQIYAVEGCVYLASCATVSEPMIKVLVD 240  
QY 240 PDKEMFLKAGGFAMIFGPDGRALAEPLPETEGLLVADIDLGMIALAKAAADPAGHYSR 299  
Db 241 DEQRKLIRGGGFARIIIGPDGRDIATPLADEBEGILYADIDLSAITLAKQAADPVGHYSR 300  
QY 300 PDVTRLLDRRPAQRVVT 317  
Db 301 PDVLSLNFNRHTTTPVNT 318

## RESULT 3

A43470  
aliphatic nitrlase - Rhodococcus rhodochrous  
C/Species: Rhodococcus rhodochrous  
C/Date: 10-Jun-1993 #sequence\_revision 18-Nov-1994 #text\_change 31-Dec-2004  
C/Accession: A43470  
R/Kobayashi, M.; Yanaka, N.; Nagasawa, T.; Yamada, H.  
Biochemistry 31, 9000-9007, 1992  
A/Title: Primary structure of an aliphatic nitrlase-degrading enzyme, aliphatic nitrlase  
A/Reference number: A43470; MUID:93003039; PMID:1390687  
A/Contents: K22  
A/Accession: A43470  
A/Status: preliminary  
A/Molecule type: DNA; protein  
A/Residues: 1-383 <KOB>  
A/Cross-references: UNIPROT:Q02068; UNIPARC:UPI0000130516; GB:D12583; NID:9216931; PIDN:  
A/Note: sequence extracted from NCBI backbone (NCBIN:114184, NCBIT:114185)  
C/Superfamily: nitrlase (carbon-nitrogen hydrolase)

Query Match 45.7%; Score 825.5; DB 2; Length 383;  
Best Local Similarity 47.2%; Pred. No. 2.6e-62;  
Matches 169; Conservative 49; Mismatches 125; Indels 15; Gaps 6;

QY 2 SEPMTRYRG---AAVQAPVFLDLDRVTEKAIGLEQAKODVRLIAPETWIPGYPMW 57  
Db 3 SNBELKTTGKVKVATVQAEVYIIDADATIDKAIQIFEEBAKNGAEFLAPFEVWIPGYPMW 62  
QY 58 IWLGAAPMGR-FVQRYFENSILVSGSKOMQALADAAARRHGMVAVAGYSEKAGSLYMGQA 116  
Db 63 AWIGDYKMAVSDPIPKYHENSILTGDDRMKRLDIAARONNIALVMWSEKDGASRYLSQV 122  
QY 117 IFPPDGLIAARRKLTPTAERTVFGSGDSHLAVHDTAIGRLALCMWHIOPLSKYAM 176  
Db 123 FIDQNGDIYANRRKLKPTHERKTYGSGNDTFLTHDFGGRVGLNCMWHEFQTLTYKAM 182  
QY 177 YAADEQHVAVSWPSFSIYRGAAYALGPEVNTAASQIYAVEGCVYLASCATVSEPMIKVL 236  
Db 183 YSLNEQHVAVSWPAMFALTPDVHQLSVEANDVTYRSYALIEGQTFVLAETHVIGKATQDLF 242  
QY 237 VDPFD-KEMFLKAGGFAMIFGPDGRALAEPLPETEGLLVADIDLGMIALAKAAADPAG 295  
Db 243 AGDDAKRALPLPGQWARIYGPDKSLAPEDDEGLLYASLDLEQIILAKAAADPAG 302  
QY 296 HYGRPDVTRLLDRR---PAQRVVTLDAAFEQNEDEKDAAPALRVVA-----ESAAA 345  
Db 303 HYSRPVLSIKIDIRNHTPVQ-YITADGRISLNSNRVENYRLHQLDIEKYEAEAA 359

## RESULT 4

A28658  
nitrlase (EC 3.5.5.1) - Klebsiella ozaenae  
C/Species: Klebsiella ozaenae  
C/Date: 10-Sep-1999 #sequence\_revision 10-Sep-1999 #text\_change 31-Dec-2004  
C/Accession: A28658  
R/Stalker, D.M.; Mal'j, L.D.; McBride, K.E.  
J. Biol. Chem. 263, 6310-6314, 1988  
A/Title: Purification and properties of a nitrlase specific for the herbicide bromoxynil  
A/Reference number: A28658; MUID:88198177; PMID:2834373  
A/Accession: A28658  
A/Molecule type: DNA  
A/Residues: 1-349 <STA>  
A/Cross-references: UNIPROT:P10045; UNIPARC:UPI000013051A; GB:J03196; NID:9149174; PIDN:  
C/Superfamily: nitrlase (carbon-nitrogen hydrolase)  
C/Keywords: hydrolase

Query Match 34.1%; Score 616; DB 1; Length 349;  
Best Local Similarity 41.1%; Pred. No. 1.4e-44;  
Matches 139; Conservative 64; Mismatches 113; Indels 22; Gaps 9;

QY 6 TKYGAAYVQAPVFLDLDRVTEKAIGLEQAKODVRLIAPETWIPGYPMWIGAPAW 65  
Db 3 TTFKAAAVQAEFPVMDAAATADTKVTTLVAKAAAGAOVLNPELMIPGYPMFLTHNGTE 62  
QY 66 GMRFVQRYFENSILVSGSKOMQALADAAARRHGMVAVAGYSEKAGSLYMGQALFEPDGLI 125  
Db 63 TLPFIKRYKQAIADGPEIKIRCAQOEHNIALSFYSEKAGRTLYMSQMLIDADGITK 122  
QY 126 AARRKLTPTAERTVFGSGDSHLAVHDTAIGRLALCMWHIOPLSKYAMVADEQYHV 185  
Db 123 IRRRLKLTPTAERTVFGSGDSDDLQVACTSVGRGALCMWENLQSLNLFALAAEEOIHI 182  
QY 186 ASWPSFSIYRGAAYALGPEVNTAASQIYAVEGCVYLASCATVSEPMIKV--LVDPDKE 243  
Db 183 SAMF-FTL--GSPVAVGDSIG-AINQVYAATGTFVLMSTGVVGTGIAAFAIEDRYNPN 238  
QY 244 MFLKAGGFAMIFGPDGRALAEPLPETEGLLVADIDLGMIALAKAAADPAGHYSRPVT 303  
Db 239 QYL--GGGVARIYGPDMQKSKSLPTEEGIVYAEIDISMEAAKYSIDPTGHYSRPDV 296  
QY 304 RLLUD--RRPAQRVVTLD-----AAFEQNEDEKGD 331  
Db 297 SVSINRQRPVASEV-IDSGNDEDERAACEP---DEGD 330

## RESULT 5

J04212

nitrilase (EC 3.5.5.1) - Comamonas testosteroni

C1Species: Comamonas testosteroni

C1Date: 14-Nov-1995 #sequence\_revision 08-Feb-1996 #text\_change 31-Dec-2004

C1Accession: J04212; PC4056

R1Levy-Schli, S.; Soubrrier, F.; Cruz-Le Coq, A.M.; Faucher, D.; Crouzet, J.; Petre, D.

Gene 161, 15-20, 1995

A1Title: Aliphatic nitrilase from a soil-isolated Comamonas testosteroni sp.: Gene clone

A1Reference number: J04212; MUID:95369726; PMID:7642130

A1Accession: J04212

A1Molecule type: DNA

A1Residues: 1-354 &lt;LE&gt;

A1Crosss-references: UNIPROT:O59329; UNIPARC:UPI00000988BB; GB:L32589; NID:g1082008; PIDN

A1Accession: PC4056

A1Molecule type: protein

A1Residues: 154-162/295-302/323-340 &lt;LE2&gt;

A1Crosss-references: UNIPARC:UPI0000175832; UNIPARC:UPI0000175833; UNIPARC:UPI0000175834

C1Comment: This enzyme is active on adiponitrile and cyanovaleic acid.

C1Genetics:

A1Gene: nltA

C1Superfamily: nitrilase (carbon-nitrogen hydrolase)

C1Keywords: hydrolase

F163/Active site: Cys #status predicted

Query Match 32.8%; Score 592.5; DB 2; Length 354;

Best Local Similarity 39.4%; Pred. No. 1.4e-42;

Matches 137; Conservative 51; Mismatches 133; Indels 27; Gaps 7;

QY 11 AAVQAPVFLDRLTVEKALIGLEQAKQDVRLIAPETWIPGYPFWIMLG-----AAAW 65

DB 10 AAVQAPVFLDRLTVEKALIGLEQAKQDVRLIAPETWIPGYPFWIMLG-----AAAW 69

QY 66 GMRFOQRYFENSLVRGSKQWALADARRHGMHVVAGYSERAGSLYMGQALFGPDGLI 125

DB 70 AV-----LFKALIEIPSEKVEQISDAKKNQYVVSSEKONASLYTLQMLFDPNGNLI 124

QY 126 AARRKLTPTHARTVFGEGSHLAVHDTAIGRLGALCCEHIOPLSKYAMTADEQVAV 185

DB 125 GKHRFKETTSERRAVWGDDGMAVVFTEYGNLGLQCEHIALPLNIAAMGSLNEQVAV 184

QY 186 ASWPEF-----SLYRGMAVYALGPEVNTAASQIYAVGGCVYLASCATVSEPMKVLVD 238

DB 185 ASWPEFVPRKAVSSVSSVSCSTNMTQIISQFAISNQVYVIVSTNLVGDMDITGK 244

QY 239 TPDKEMFKAGGFGFMIFGPDGRALAPLPETEGELVADIDLGMIALAKAADPAGHY 298

DB 245 DEFSKNFPLPSGNTRAITISNTEILAS-IPQDAEIAVAEIDLNQIITGKMLDPAGHY 303

QY 299 RPDVTRLLDRR--PAQRV--VTIDAAFEQNDKGD--APALRV 337

DB 304 TPGFLSLTFDQSEHVPVKKIGQTNHFISEDLHEDKKMDLTPRRV 351

## RESULT 6

S77025

nitrilase (EC 3.5.5.1) - Synechocystis sp. (strain PCC 6803)

N1Alternate names: protein sll0784

C1Species: Synechocystis sp.

A1Variety: PCC 6803

C1Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 31-Dec-2004

C1Accession: S77025

R1Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N.;

O. K.; Okumura, S.; Shimo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda

DNA Res. 3, 109-116, 1996

A1Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis

S77025

A1Reference number: S74322; MUID:97061201; PMID:8905231

A1Accession: S77025

A1Status: nucleic acid sequence not shown; translation not shown

A1Molecule type: DNA

A1Residues: 1-346 &lt;KAN&gt;

A1Crosss-references: UNIPROT:O55949; UNIPARC:UPI00000D7133; EMBL:D64005; GB:AB001339; NID

A1Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996

C1Genetics:

A1Start codon: GTG

C1Superfamily: nitrilase (carbon-nitrogen hydrolase)

C1Keywords: hydrolase

Query Match 32.4%; Score 584.5; DB 2; Length 346;

Best Local Similarity 42.1%; Pred. No. 6.5e-42;

Matches 128; Conservative 53; Mismatches 114; Indels 9; Gaps 6;

QY 9 RGAAPVFLDRLTVEKALIGLEQAKQDVRLIAPETWIPGYPFWIMLGAP-VGM 67

DB 14 RGAAPVFLDRLTVEKALIGLEQAKQDVRLIAPETWIPGYPFWIMLGAP-VGM 73

QY 68 RFOQRYFENSLVRGSKQWALADARRHGMHVVAGYSERAGSLYMGQALFGPDGLIA 127

DB 74 RFLKQGAATVPG-KVQAIAQAKTGMMVVGVNREBGSLYNTQLIFDADALVTK 132

QY 128 RRLKLTPTHARTVFGEGSHLAVHDTAIGRLGALCCEHIOPLSKYAMTADEQVAV 187

DB 133 RRLKLTPTHARTVFGEGSHLAVHDTAIGRLGALCCEHIOPLSKYAMTADEQVAV 192

QY 188 WPSFLYRGMAVYALGPEVNTAASQIYAVGGCVYLASCATVSEPMKVLVDTPDKEMFLK 247

DB 193 WPSFLYRGMAVYALGPEVNTAASQIYAVGGCVYLASCATVSEPMKVLVDTPDKEMFLK 246

QY 248 AGGFGFMIFGPDGRALAPLPETEGELVADIDLGMIALAKAADPAGHYRDPVTRLL 307

DB 247 AGGFGFMIFGPDGRALAPLPETEGELVADIDLGMIALAKAADPAGHYRDPVTRLL 305

QY 308 DRP 311

DB 306 NNOP 309

## RESULT 7

J01613

cyanide hydratase (EC 4.2.1.66) - imperfect fungus (Gloeocercospora sorghi)

C1Species: Gloeocercospora sorghi

C1Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 31-Dec-2004

C1Accession: J01613

R1Wang, P.; VanEtten, H.D.

Biochem. Biophys. Res. Commun. 187, 1048-1054, 1992

A1Title: Cloning and properties of a cyanide hydratase gene from the phytopathogenic fun

A1Reference number: J01613; MUID:92412068; PMID:1382413

A1Accession: J01613

A1Molecule type: DNA

A1Residues: 1-368 &lt;MAN&gt;

A1Crosss-references: UNIPROT:P32964; UNIPARC:UPI0000128C2E; GB:M99044; NID:g168196; PIDN

C1Comment: This enzyme converts HCN to formamide.

C1Genetics:

A1Gene: chC

C1Superfamily: nitrilase (carbon-nitrogen hydrolase)

C1Keywords: carbon-oxygen lyase; glycoprotein; hydro-lyase

F1286/Binding site: carbonyl (Asn) (covalent) #status predicted

Query Match 31.1%; Score 562; DB 2; Length 368;

Best Local Similarity 40.7%; Pred. No. 5.7e-40;

Matches 132; Conservative 50; Mismatches 124; Indels 18; Gaps 7;

QY 4 PMTYRGAAPVFLDRLTVEKALIGLEQAKQDVRLIAPETWIPGYPFWIMLGAP 63

DB 2 PINKYKAAVVTSEFWENLEGGVKTIEFIEAGKACKLIAPEVWIPGYPWMMKVNY 61

QY 64 AMGRFOQRYFENSLVRGSKQWALADARRHGMHVVAGYSERAGSLYMGQALFGPDGL 123

DB 62 LQSPMPKAKRENSIDAMSSERRIRPAARNQIYVIGSEIDHATLYTLQVILISPLGD 121

QY 124 LIAARRKLTPTHARTVFGEGSHLAVHDTAIGRLGALCCEHIOPLSKYAMTADEQ 182

DB 122 VINRRKIKPTTHARTVFGEGSHLAVHDTAIGRLGALCCEHIOPLSKYAMTADEQ 181

183 VHVASWPSFSLYRGMAVALGPEVNT-----AASQI-----YAVEGGCYTLASCATVSPENIK 234  
Db 182 IHVAAWP-----VYPDLSTQKVHPDPATNVADPASDLVTPAYAIETGTWVLAPQRISEGLK 238  
Qy 235 VLVDP-----DKEMFLKAGGFAMITGPGRALAEPLETEBGLVADIDLCIMIALAKAA 291  
Db 239 --RHPPEGVPEPTDATPTYNHARIRFPDGSLYAKPAVDF-DGLMIVYDIDLESHTLKALA 295  
Qy 292 DPAGHYSRPDVTRLILDRRPAQRV 315  
Db 296 DPAGHYMRPDLIRLLVDTRRKELV 319

## RESULT 8

T49148  
nitrilase (EC 3.5.5.1) 3 [imported] - Arabidopsis thaliana  
N:Alternate names: protein T10D17.110  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C>Date: 02-Jun-2000 #sequence revision 02-Jun-2000 #text\_change 31-Dec-2004  
C/Accession: T49148; T52261; T52264  
R/D:Angelo, M.; Vezzi, A.; Modesto, D.; Pigazzi, M.; Valle, G.; Mewes, H.W.; Rudd, S.; L submitted to the Protein Sequence Database, April 2000  
A:Reference number: 225017  
A:Accession: T49148  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-346 <DAN>  
A:Cross-references: UNIPROT:P46010; UNIPARC:UPI000000BDB6; EMBL:AL353865; GSPDB:GN00061;  
A:Experimental source: cultivar Columbia; BAC clone T10D17  
R/Hillebrand, H.; Bartling, D.; Weiler, E.W.  
Plant Mol. Biol. 36, 89-99, 1998  
A:Title: Structural analysis of the nit2/nit1/nit3 gene cluster encoding nitrilases, enz  
A:Reference number: 226007; MUID:98145459; PMID:9484465  
A:Accession: T52261  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-346 <HIL>  
A:Cross-references: UNIPARC:UPI000000BDB6; EMBL:Y07648; PIDN:CAA68936.2  
A:Experimental source: cultivar Columbia  
R/Bartel, B.; Fink, G.R.  
Proc. Natl. Acad. Sci. U.S.A. 91, 6649-53, 1994  
A:Title: Differential regulation of an auxin-producing nitrilase gene family in Arabidop  
A:Reference number: 224515; MUID:8022831; PMID:8022831  
A:Accession: T52264  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-346 <BAR>  
A:Cross-references: UNIPARC:UPI000000BDB6; EMBL:U09959; PIDN:AAA19627.1  
A:Experimental source: cultivar Columbia  
C/Genetics:  
A:Gene: ATSP:T10D17.110; nit3  
A:Map position: 3  
A:Introns: 44/1; 104/1; 202/1; 296/1  
C:Superfamily: nitrilase (carbon-nitrogen hydrolase)  
C:Keywords: hydrolase

Query Match 25.9%; Score 468.5; DB 2; Length 346;  
Best Local Similarity 33.0%; Pred. No. 4.5e-32;  
Matches 115; Conservative 65; Mismatches 121; Indels 47; Gaps 9;

Qy 3 EPMTRYGAAVQAQAVFLDLDRIVEKAIGLIEQAQKQDVRLLIAPETWIPGYPMWIGA 62  
Db 20 DPSSIVRTIVQSSIVYNDTPATIDKAEKIVEAASKAGKLVLPFEARIGGVP----- 72  
Qy 63 PAMGRF-----VQRYFENSIVRGSKQWQALADARRHGMHNVAGSERAG 108  
Db 73 --RGRFGLAVGVHNEBGRDEFRRYHSAIIVPGPEVARLADVARKNVYLVMGAIEKGY 130  
Qy 109 SLVWGQAIFGPDDGLIARRKLTPTAERTVFGSGDSHLAVHDTAIGRLGALCWEHI 168  
Db 131 YTVCTALFFSPQGFLOKHKRWKMTSLERCIMWGDSSTIPVYDTPIGKIGALCWEHR 190

Qy 169 QPLSKVMAVADBEQVHVASWPSFSLYRGMAVALGPEVNTAASQIY-AVEGGCYTLAS--- 224  
Db 191 MPLYRTALVYKGLIEIYCAPADVSL-----EMQASMIHIVEGGCYTLASHQF 238  
Qy 225 CATVS-PE-----MKVVDTPDKEMFLKAGGFAMITGPGRALAEPLETEBGLVADI 279  
Db 239 CKRREFPEHPDYLVNDIVDTKEHDPVYSGG--SVIISPLGKVLAGFNYS--BGLVADL 295  
Qy 280 DLGMIALAKAAADPAGHYSRPDVTRLILDRRPAQRVTVTLDAEPONE 327  
Db 296 DLGDIAKALVYFDVVGHYSKPDIFNLTVNEHPKPVTFMTKVEKADE 343

## RESULT 9

T49147  
nitrilase (EC 3.5.5.1) 1 [imported] - Arabidopsis thaliana  
N:Alternate names: protein T10D17.100  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C>Date: 02-Jun-2000 #sequence revision 02-Jun-2000 #text\_change 31-Dec-2004  
C/Accession: T49147; T52260; T52263  
R/D:Angelo, M.; Vezzi, A.; Modesto, D.; Pigazzi, M.; Valle, G.; Mewes, H.W.; Rudd, S.; L submitted to the Protein Sequence Database, April 2000  
A:Reference number: 225017  
A:Accession: T49147  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-346 <DAN>  
A:Cross-references: UNIPROT:P39661; UNIPARC:UPI0000001AE2; EMBL:AL353865; GSPDB:GN00061;  
A:Experimental source: cultivar Columbia; BAC clone T10D17  
R/Hillebrand, H.; Bartling, D.; Weiler, E.W.  
Plant Mol. Biol. 36, 89-99, 1998  
A:Title: Structural analysis of the nit2/nit1/nit3 gene cluster encoding nitrilases, enz  
A:Reference number: 226007; MUID:98145459; PMID:9484465  
A:Accession: T52260  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-346 <HIL>  
A:Cross-references: UNIPARC:UPI0000001AE2; EMBL:Y07648; PIDN:CAA68935.2  
A:Experimental source: cultivar Columbia  
R/Bartel, B.; Fink, G.R.  
Proc. Natl. Acad. Sci. U.S.A. 91, 6649-53, 1994  
A:Title: Differential regulation of an auxin-producing nitrilase gene family in Arabidop  
A:Reference number: 224515; MUID:8022831; PMID:8022831  
A:Accession: T52263  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-346 <BAR>  
A:Cross-references: UNIPARC:UPI0000001AE2; EMBL:U38845; PIDN:AAB05221.1  
A:Experimental source: cultivar Columbia  
C/Genetics:  
A:Gene: ATSP:T10D17.100; nit1; NIT1  
A:Map position: 3  
A:Introns: 44/1; 104/1; 202/1; 296/1  
C:Superfamily: nitrilase (carbon-nitrogen hydrolase)  
C:Keywords: hydrolase

Query Match 25.9%; Score 467; DB 2; Length 346;  
Best Local Similarity 34.1%; Pred. No. 6.1e-32;  
Matches 121; Conservative 50; Mismatches 122; Indels 62; Gaps 8;

Qy 4 EPMTRYGAAVQAQAVFLDLDRIVEKAIGLIEQAQKQDVRLLIAPETWIPGYPMWIGA 63  
Db 21 PSTIVRTIVQSSIVYNDTPATIDKAEKIVEAASKAGKLVLPFEARIGGVP----- 72  
Qy 64 AMGRF-----VQRYFENSIVRGSKQWQALADARRHGMHNVAGSERAG 109  
Db 73 --RGRFGLAVGVHNEBGRDEFRRYHSAIIVPGPEVARLADVARKNVYLVMGAIEKGY 131  
Qy 110 SLVWGQAIFGPDDGLIARRKLTPTAERTVFGSGDSHLAVHDTAIGRLGALCWEHI 169  
Db 132 YTVCTALFFSPQGFLOKHKRWKMTSLERCIMWGDSSTIPVYDTPIGKIGALCWEHR 191  
Qy 170 PLSKVMAVADBEQVHVA-----SWPSFSLYRGMAVALGPEVNTAASQIYAVEGGCYTL 222



```
DB 192 PLVRLALAKGELIYCAPADSGKEMQSMHL-----IAIEGCGFVL 233
OY 223 ASCATVSEPMIKVLTDP-----DKEMFLKAGGPFAMIFGPDGRALAEPLPETEERG 273
DB 224 SACQPCOR---KHFPDHDYLTFTDWDYDKEDHSIVSQGGSVILISPLGVLADPNFES-EG 289
OY 274 LTVADIDIGMTALAKAADPAGHYSRPDVTRLDDRPAQRVYTLDAAPFQONED 328
DB 290 LVTADIDIGDIAKARALKYFDSVGHYSRPVLTHTVNEHP-RKSVTFVTKEKXEDD 343

RESULT 10
S31969
nitrilase (EC 3.5.5.1) [similarity] - Arabidopsis thaliana
N/Alternate names: T10D17.90
C/Species: Arabidopsis thaliana (mouse-ear cress)
C/Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 31-Dec-2004
C/Accession: S31969; T49146; T52258
R/Bartling, D.; Seedorf, M.; Schmidt, R.C.; Weiler, E.W.
submitted to the EMBL Data Library, September 1992
A/Description: Arabidopsis thaliana nitrilases are encoded by a two-member gene family:
A/Reference number: S31969
A/Accession: S31969
A/Molecule type: mRNA
A/Residues: 1-339 <BAR>
A/Cross-references: UNIPROT:P32962; UNIPARC:UPI0000001444; EMBL:X68305; NID:g22655; PID:
R/D'Angelo, M.; Vezzi, A.; Modesto, D.; Pigazzi, M.; Valle, G.; Mewes, H.W.; Rudd, S.; I
submitted to the Protein Sequence Database, April 2000
A/Reference number: 225017
A/Accession: T49146
A/Status: preliminary
A/Molecule type: DNA
A/Residues: 1-339 <DNA>
A/Cross-references: UNIPARC:UPI0000001444; EMBL:AL353865; GSPDB:GN00061; ATSP:T10D17.90
R/Bartling, D.; Seedorf, M.; Schmidt, R.C.; Weiler, E.W.
Proc. Natl. Acad. Sci. U.S.A. 91, 6021-5, 1994
A/Title: Molecular characterization of two cloned nitrilases from Arabidopsis thaliana:
A/Reference number: 224514; MUID:8016109; PMID:8016109
A/Accession: T52258
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: mRNA
A/Residues: 1-339 <BAR>
A/Cross-references: UNIPARC:UPI0000001444; EMBL:X68305; PIDN:CAA48377.1
A/Experimental source: cultivar Landsberg erecta
C/Genetics:
A/Gene: ATSP:T10D17.90
A/Map position: 3
A/Introns: 37/1; 97/1; 195/1; 289/1
C/Superfamily: nitrilase (carbon-nitrogen hydrolase)
C/Keywords: hydrolase

Query Match 25.7%; Score 464; DB 2; Length 339;
Best Local Similarity 34.6%; Pred. No. 1.1e-31;
Matches 119; Conservative 54; Mismatches 127; Indels 44; Gaps 8;

OY 6 TKYGAAGVAAVFLDLDRTEKAIGLEQAKQDVRLIAPETWIPGP--FWIWLGA 63
DB 16 TIVATIVQASTVNDTPATLEKANKFIVEAASKSELVPEPAFIGYPRGFRGLGVG 75
OY 64 AW---GKRFVORYFENSLVRGSKQWALADARRHGHHVAVGYSEBAGSLYMGQATFGP 120
DB 76 VHNREGDEPRKHYASAIKVPGEVEKLAELAGKNNVYLWGALIEKDGTYLTCTALFFSP 135
OY 121 DGDIIARRKILKPTAERTVFGEGSHLAVHDPTAIGRLGALCWEHIQPLSKYAMVAD 180
DB 136 QGQFLGKRRKLMPTSLERCIWGGDGSITIPYDPIGKGAALCWNEMPLYRTALYAKG 195
OY 181 EQVHVA-----SWPSFLYRGMAVYALGPEVNTAAQIYAVEGGCYTLASCATVSPEMI 233
DB 196 IELYCAPADSGKEMQSMHL-----IAIEGCGFVLSACQPC---LR 234
```

```
OY 234 KVLVDP-----DKEMFLKAGGPFAMIFGPDGRALAEPLPETEGLVADIDLGM 284
DB 235 KQFDHPDYLTFTDWDYDKEDPSIVSQGGSVILISPLGVLADPNFES-EGLTITADLDGDV 293
OY 285 ALAKAADPAGHYSRPDVTRLDDRPAQRVYTLDAAPFQONED 328
DB 294 ARAKLYFDSVGHYSRPVLTHTVNEHP-KKVTFTISKVEKXEDD 336

RESULT 11
T52259
nitrilase (EC 3.5.5.1) 2 [imported] - Arabidopsis thaliana
C/Species: Arabidopsis thaliana (mouse-ear cress)
C/Date: 20-Oct-2000 #sequence_revision 20-Oct-2000 #text_change 31-Dec-2004
C/Accession: T52259
R/Hillebrand, H.; Bartling, D.; Weiler, E.W.
Plant Mol. Biol. 36, 89-99, 1998
A/Title: Structural analysis of the nlt2/nlt1/nlt3 gene cluster encoding nitrilases, enz
A/Reference number: 226007; MUID:98145459; PMID:9484465
A/Accession: T52259
A/Status: preliminary; translated from GB/EMBL/DBJ
A/Molecule type: DNA
A/Residues: 1-339 <HLI>
A/Cross-references: UNIPROT:O04907; UNIPARC:UPI00000A5899; EMBL:Y07648; PIDN:CAA68934.3
A/Experimental source: cultivar Columbia
C/Genetics:
A/Gene: nlt2
A/Introns: 37/1; 97/1; 195/1; 289/1
C/Superfamily: nitrilase (carbon-nitrogen hydrolase)
C/Keywords: hydrolase

Query Match 25.6%; Score 463; DB 2; Length 339;
Best Local Similarity 34.6%; Pred. No. 1.3e-31;
Matches 119; Conservative 54; Mismatches 127; Indels 44; Gaps 8;

OY 6 TKYGAAGVAAVFLDLDRTEKAIGLEQAKQDVRLIAPETWIPGP--FWIWLGA 63
DB 16 TIVATIVQASTVNDTPATLEKANKFIVEAASKSELVPEPAFIGYPRGFRGLGVG 75
OY 64 AW---GKRFVORYFENSLVRGSKQWALADARRHGHHVAVGYSEBAGSLYMGQATFGP 120
DB 76 VHNREGDEPRKHYASAIKVPGEVEKLAELAGKNNVYLWGALIEKDGTYLTCTALFFSP 135
OY 121 DGDIIARRKILKPTAERTVFGEGSHLAVHDPTAIGRLGALCWEHIQPLSKYAMVAD 180
DB 136 QGQFLGKRRKLMPTSLERCIWGGDGSITIPYDPIGKGAALCWNEMPLYRTALYAKG 195
OY 181 EQVHVA-----SWPSFLYRGMAVYALGPEVNTAAQIYAVEGGCYTLASCATVSPEMI 233
DB 196 IELYCAPADSGKEMQSMHL-----IAIEGCGFVLSACQPC---LR 234
OY 234 KVLVDP-----DKEMFLKAGGPFAMIFGPDGRALAEPLPETEGLVADIDLGM 284
DB 235 KQFDHPDYLTFTDWDYDKEDPSIVSQGGSVILISPLGVLADPNFES-EGLTITADLDGDV 293
OY 285 ALAKAADPAGHYSRPDVTRLDDRPAQRVYTLDAAPFQONED 328
DB 294 ARAKLYFDSVGHYSRPVLTHTVNEHP-KKVTFTISKVEKXEDD 336

RESULT 12
S22398
nitrilase (EC 3.5.5.1) - Arabidopsis thaliana
C/Species: Arabidopsis thaliana (mouse-ear cress)
C/Date: 13-Jan-1995 #sequence_revision 13-Jan-1995 #text_change 31-Dec-2004
C/Accession: S22398
R/Bartling, D.; Seedorf, M.; Mitchofer, A.; Weiler, E.W.
Eur. J. Biochem. 205, 417-424, 1992
A/Title: Cloning and expression of an Arabidopsis nitrilase which can convert indole-3-a
A/Reference number: S22398; MUID:92209532; PMID:1555601
A/Accession: S22398
A/Molecule type: mRNA
A/Residues: 1-346 <BAR>
```

A:Cross-references: UNIPROT:P32961; UNIPARC:UP10000130513; EMBL:X63445; NID:g16399; PIDN  
C:Superfamily: nitrilase (carbon-nitrogen hydrolase)  
C:Keywords: hydrolase

Query Match 25.5%; Score 461; DB 2; Length 346;  
Best Local Similarity 33.8%; Pred. No. 2e-31;  
Matches 120; Conservative 51; Mismatches 122; Indels 62; Gaps 8;

QY 4 PMTKRGAAVQAAPVFLDLDRTEKAIIGLIEQAKODVRLIAPETWIPGYPFMIWLGAP 63  
DB PSTIVRTVIVQSGSTVYNDTPATIDKAEKYIVEAASKGAELVLPFGGIGYF----- 72  
QY 64 AMGRF-----VQRYFENSLVSGSKOMQLADARRHGHHVAGYSERAGG 109  
DB 73 -RGFRFLGAVGVHNEGRDEFKRYHSAIHVGPPEVARLAVARNHYLVGAIKEKGY 131  
QY 110 SLVNGQALFGPDGDLIARRKLTPTHAERTVFGSGDGLAVHDTAIGRLGALCWEHIQ 169  
DB 132 TLCTVLFPSFGQFLGKHKRLMPTSLERCIMGGDSITIPYDTPIGKGLACWENRM 191  
QY 170 PLSTKAMTADEQVHVA-----SWPSFSLYRGMAYALGPEVNTAASQIYAVGGCYVL 222  
DB 192 PLYRTALYAKGIELYCAPTADGSKWQSSMLH-----IAIEGGCFVL 233  
QY 223 ASCATVSPKIKVLDTP-----DKEMFLKAGGFAMIFGPDGRALAPLPETPEG 273  
DB 234 SAGCFQR---KHPPDHDLFTDMYDKHDSIVSGGSVITISPLGVLAGPNEF-EG 289  
QY 274 LVVADIDIGMALAKAADPAGHYSRPVTLLDRRPAQRVVTIDAAEPONED 328  
DB 290 LVTDIDIGDIARAKLYPDSVGYSRPDLHLTVENHP-RKSVTFVTKVKAEDD 343

## RESULT 13

TS2262  
nitrilase (EC 3.5.5.1) 2 (imported) - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C>Date: 20-Oct-2000 #sequence\_revision 20-Oct-2000 #text\_change 31-Dec-2004  
C:Accession: T52262  
R:Bartel, B.; Fink, G.R.  
Proc. Natl. Acad. Sci. U.S.A. 91, 6649-53, 1994  
A:Title: Differential regulation of an auxin-producing nitrilase gene family in Arabidop  
A:Reference number: Z44515; MUID:8022831; PMID:8022831  
A:Accession: T52262  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-339 <BAR>  
A:Cross-references: UNIPROT:P32962; UNIPARC:UP1000016DA4C; EMBL:U38845; PIDN:AAB05220.1  
A:Experimental source: cultivar Columbia  
C:Genetics:  
A:Gene: NIT2  
A:Map position: 3  
A:Insertions: 37/2; 97/1; 195/1; 289/1  
C:Superfamily: nitrilase (carbon-nitrogen hydrolase)  
C:Keywords: hydrolase

Query Match 25.2%; Score 456; DB 2; Length 339;  
Best Local Similarity 34.3%; Pred. No. 5.1e-31;  
Matches 118; Conservative 54; Mismatches 128; Indels 44; Gaps 8;

QY 6 TKYGAAVQAAPVFLDLDRTEKAIIGLIEQAKODVRLIAPETWIPGYPFMIWLGAP 63  
DB 16 TIVATVQASTVNDTPATIDKANKPIVEATKSELVPEPAFIGYPRGFRGQVG 75  
QY 64 AM---GKRFVQRYEENSLVRSGKQMALADARRHGHHVAVGYSERAGGSLVNGQATGCP 120  
DB 76 VHNNEGRDEFKRYHSAIKVGPPEVKLAELAGKNNVYLWGALEKQDGYTLYCYALFPSP 135  
QY 121 DGDILIAARRKLTPTHAERTVFGSGDGLAVHDTAIGRLGALCWEHIQPLSTKAMTAAD 180  
DB 136 QGQFLGKHKRLMPTSLERCIMGGDSITIPYDTPIGLGAALCWEENMPLRYALYAKG 195  
QY 181 EQVHVA-----SWPSFSLYRGMAYALGPEVNTAASQIYAVGGCYVLASCATVSPMI 233

DB 196 IELYCAPTADGSKWQSSMLH-----IAIEGGCFVLSSACQPC---LR 234  
QY 234 KVLVDTP-----DKEMFLKAGGFAMIFGPDGRALAPLPETEGGLVADIDIGMI 284  
DB 235 KDFPDHDLFTDMYDKHDSIVSGGSVITISPLGVLAGPNEF-EGLTITADLDIGDV 293

QY 285 ALAKAADPAGHYSRPVTLLDRRPAQRVVTIDAAEPONED 328  
DB 294 ARAKLYPDSVGHYSRPDLHLTVENHP-KRPVTFISKVEKAEDD 336

## RESULT 14

TS2266  
nitrilase-like protein (imported) - rice  
C:Species: Oryza sativa (rice)  
C>Date: 20-Oct-2000 #sequence\_revision 20-Oct-2000 #text\_change 31-Dec-2004  
C:Accession: T52266  
R:Chiba, R.; Dohmoto, M.; Yamaguchi, K.  
submitted to the EMBL Data Library, May 1999  
A:Description: Oryza sativa, a gene for nitrilase-like protein.  
A:Reference number: Z26008  
A:Accession: T52266  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA  
A:Residues: 1-362 <CHI>  
A:Cross-references: UNIPROT:Q9SXK6; UNIPARC:UP1000009C92A; EMBL:AB027054; PIDN:BA077679.  
A:Experimental source: cultivar Nipponbare  
C:Genetics:  
A:Gene: ONIT4  
C:Superfamily: nitrilase (carbon-nitrogen hydrolase)

Query Match 23.3%; Score 420; DB 2; Length 362;  
Best Local Similarity 34.1%; Pred. No. 6.3e-28;  
Matches 117; Conservative 57; Mismatches 137; Indels 32; Gaps 10;

QY 9 RGAAVQAAPVFLDLDRTEKAIIGLIEQAKODVRLIAPETWIPGYPFMIWLGAP 60  
DB 32 RATVQASTVFEYDTPATIDKAEKRIEBAAGYSQLVFPEPAFVGSGYRGSTFGGANISI 91  
QY 61 GAPA-WGMRFYQRYFENSLVRSGKQMALADARRHGHHVAVGYSERAGGSLVNGQALIFG 119  
DB 92 GNPDKKKEERKTHAAIIEVPGPEVRLAMAKYVFLVMGYTEREGTILYCSULFDF 151  
QY 120 PDGDLIAARRKLTPTHAERTVFGSGDGLAVHDTAIGRLGALCWEHIQPLSTKAMTAAD 179  
DB 152 PLGRLYLGKHKRLMPTSLERCIMGGDSITIPYDTPIGKGLACWENKMPLELTALYK 211  
QY 180 DEQVHVASWPSFSLYRGMAYALGPEVNTAASQIYAVGGCYVLAS---CATVS---PEM 232  
DB 212 GIELYCA--PT-----ADSRQVQASWTHIALGEGCFVLSSANQFCRRKDYPPPEY 260  
QY 233 IKV-LVDTPPKEMFLKAGGFAMIFGPDGRALAPLPETEGGLVADIDIGMALAKAA 291  
DB 261 VFTGLGEBSPDVTVCQG--SVIISPSGEVLAPNVE-GEALITADLDIGEIVRAKDF 317  
QY 292 DPAGHYSRPVTLLDRRPAQRVVTIDAAEPONEDGAP 334  
DB 318 DVVGHYARPEVLSLVNDQP-HLPVSPTSAAEKTTAAKSDSTA 359

## RESULT 15

T03739  
nitrilase (EC 3.5.5.1) 4B - common tobacco  
C:Species: Nicotiana tabacum (common tobacco)  
C>Date: 24-Mar-1999 #sequence\_revision 24-Mar-1999 #text\_change 31-Dec-2004  
C:Accession: T03739  
R:Tsunoda, H.  
submitted to the EMBL Data Library, January 1996  
A:Reference number: Z15044  
A:Accession: T03739  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: mRNA

A;Residues: 1-348 &lt;TSU&gt;

A; Cross-references: UNIPROT:Q42966; UNIPARC:UPI000009F332; EMBL:D83078

**A; Experimental source: strain SR1; tissue-type leaf**

C;Genetics:

A;Gene: TNIT4B  
C:Superfamily:

C:Superfamily: nitrilase (carbon-nitrogen hydrolase)  
C:Keywords: hydrolase

C;keywords: nyarolase

Query Match	22.8%;	Score 411;	DB 2;	Length 348;
-------------	--------	------------	-------	-------------

Best Local Similarity 35.1%; Pred. No. 3.5e-27;  
Matches 110; Conservative EF; Mismatches 118

Matches 110; Conservative 55; Mismatches 118; Indels 30; Gaps 10;

OY	9	RGAAYQAAPVDFDLDRVTAKAIGLEIQAKODVRLLAFETWIPGP----	PWTWG-AP	63
Db	30	RATVVQASITIFDPTBATIDKABRLLAAEASYAGQLVFPEAFICGYP	PRGSTFVSIGNRT	89
OY	64	AMGMRFVQRYFENSILVRGSKOMQALADAAARRHGMMHVAGYSERAGSLYMCAI	PGPDGD	123
Db	90	AKGKEFFRYTHASALDIVPGEVDRLRAAMAGKKVYLIMVEITBEDGYTLCTVL	FPSOGH	149
OY	124	LIAARKLKTPTAERTVFGEDGGSHLVADTIAIGRLGALCWEHI	PLSKRYMYADEOV	183
Db	150	YLGRKRKLMPFLBELLIMGFDGSGTTPVYDTPLGKI GAALCMENRRPFLRTAMYAGIEI		209
OY	184	HVASPFSSLYRGMAIYALCEPVNTAASQIYAIVEGCYTLAS---	CATVS---PEMIKVY	236
Db	210	YCA--PT-----ADSRDVQWQSMITHIALEGCGFVLSANOFCRRKDYP	PREPEVVFSG	258
OY	237	VD--TPDKEMFLKAGGFMFIGPBGORALAEPLETREGLIVADIDLGMITALAKAADBA		294
Db	259	TEDLPDS--IYCAGS--SVIIISPQAVLAGNYE-GEALLISADDLDGIARAARPPFDV		313
OY	295	GHSRSPDYTRLILL	307	
Db	314	GHYARPEVLSTIV	326	

Search completed: April 27, 2006, 01:01:07  
Job time : 27.336 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: April 27, 2006, 00:50:35 ; Search time 155.523 Seconds  
(without alignments)  
1569.628 Million cell updates/sec

Title: US-09-751-299-2

Perfect score: 1806

Sequence: 1 MSEPMTKTRGAIVQAAPVFL.....EDKGDAPALRVVAESAAAAQ 346

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Uniprot 05.80:\*

1: uniprot\_sprot:\*

2: uniprot\_trembl:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1806	100.0	346	2	Q6RWM5_92222
2	1464	81.1	341	2	Q6RWR4_92222
3	1209	66.9	348	2	Q6RWS2_92222
4	1165	64.5	353	2	Q6RWF9_92222
5	1155	64.0	353	2	Q6RWF0_92222
6	1132.5	62.7	358	2	Q6RWF0_92222
7	1130	62.6	354	2	Q6RWM6_92222
8	1127	62.4	331	2	Q706Q8_P88PU
9	1125	62.3	345	2	Q6RWS3_92222
10	1099.5	60.9	334	2	Q6RWE3_92222
11	1076.5	59.6	337	2	Q6RWN9_92222
12	1073.5	59.4	338	2	Q6RWM6_92222
13	1070	59.2	336	2	Q6RWF6_92222
14	1068.5	59.2	335	2	Q6RWF7_92222
15	1068.5	59.2	336	2	Q6RWF8_92222
16	1066.5	58.9	337	2	Q6RWN8_92222
17	1064.5	58.9	338	2	Q6RWN7_92222
18	1064.5	58.9	350	2	Q5E6G1_P88FL
19	1057.5	58.6	352	2	Q6RWH0_92222
20	1052.5	58.3	338	2	Q6RWE4_92222
21	1044.5	57.8	337	2	Q6RWS3_92222
22	1043.5	57.8	338	2	Q6RWM2_92222
23	1041	57.6	336	2	Q6RWM4_92222
24	1038.5	57.5	338	2	Q6RWM5_92222
25	1038.5	57.5	338	2	Q6RWM6_92222
26	1028.5	56.9	337	2	Q6RWF7_92222
27	1028.5	56.8	338	2	Q6RWF1_92222
28	1025.5	56.6	336	2	Q500U1_P88SY
29	1023	56.6	337	2	Q6RWM3_92222
30	1023	56.6	337	2	Q6RWM7_92222
31	1020.5	56.5	327	2	Q6RWH9_92222

32	1012	56.0	347	2	Q6RBS3_P88SM
33	1008	55.8	349	2	Q6RWS0_92222
34	1007	55.8	337	2	Q6RWE0_92222
35	1004.5	55.6	337	2	Q6RWM3_92222
36	986	54.6	337	2	Q6RWN2_92222
37	974.5	54.0	337	2	Q6RWN9_92222
38	972	53.8	334	2	Q6RWM3_92222
39	964	53.4	334	2	Q6RWS3_P88AJA
40	949.5	52.6	346	2	Q6RWM1_92222
41	949	52.5	333	2	Q6RWF0_92222
42	945	52.3	332	2	Q6RWM0_92222
43	869	48.1	356	1	NRLA_ALCPA
44	829.5	45.9	365	1	NRL2_RHORN
45	825.5	45.7	383	1	NRL1_RHORN

## ALIGNMENTS

Result 1					
ID	Q6RWM5_92222	PRELIMINARY;	PRT;	346 AA.	
AC	Q6RWM5;				
DT	05-JUL-2004 (TREMBLrel. 27, Created)				
DT	05-JUL-2004 (TREMBLrel. 27, Last sequence update)				
DE	05-JUL-2004 (TREMBLrel. 27, Last annotation update)				
DE	Nitric oxide (EC 3.5.5.7).				
GN	ORFNames=BD5086;				
OS	uncultured organism.				
OC	unclassified; environmental samples.				
OX	NCBI_Taxid=155900;				
RN	[1]				
RP	NUCLEOTIDE SEQUENCE.				
RX	PubMed=15066841; DOI=10.1128/ASM.70.4.2429-2436.2004;				
RA	Robertson D.E., Chaplin J.A., Desantis G., Podar M., Madden M.,				
RA	Chi E., Richardson T., Milan A., Miller M., Weiner D.P., Wong K.,				
RA	McQuaid J., Farwell B., Preston L.A., Tan X., Sneed M.A., Keller M.,				
RA	Mathur E., Kretz P.L., Burk M.J., Short J.M.,				
RT	"Exploring nitric oxide sequence space for enantioselective catalysis."				
RT	Appl. Environ. Microbiol. 70:2429-2436(2004).				
DR	EMBL; AY487497; AAR97444.1; -; Genomic DNA.				
DR	GO; GO:0016810; F:hydrolase activity; IEA.				
DR	GO; GO:0006807; P:nitrogen compound metabolism; IEA.				
DR	InterPro; IPR00132; N/CN hydrolase.				
DR	InterPro; IPR00132; N/CN hydrolase.				
DR	Pfam; PF00795; CN hydrolase; 1.				
DR	PROSITE; PS00263; CN HYDROLASE; 1.				
DR	PROSITE; PS00263; CN HYDROLASE; 1.				
KW	Hydrolase. 346 AA; 37432 MW; 12PE01C35ND463CE CRC64;				
SO	SEQUENCE				
Query Match					
Best Local Similarity 100.0%; Score 1806; DB 2; Length 346;					
Matches 346; Conservative 0; Mismatches 0; Indels 0; Gaps 0;					
QY	1	MSEPMTKTRGAIVQAAPVFLDLDRTEKALIGLEDAKQDRLAFPEPTWIPGFYFWL	60		
DB	1	MSEPMTKTRGAIVQAAPVFLDLDRTEKALIGLEDAKQDRLAFPEPTWIPGFYFWL	60		
QY	61	GAPAGMFPVQRYFNSLVRSKQWQALADARRHGMVAVGYSRAGSLYMGQATGP	120		
DB	61	GAPAGMFPVQRYFNSLVRSKQWQALADARRHGMVAVGYSRAGSLYMGQATGP	120		
QY	121	DGDLIAARRKLPKTAERTVFGEGDSHIAVHTAIGLALCCWEHIQPLSKYMYAD	180		
DB	121	DGDLIAARRKLPKTAERTVFGEGDSHIAVHTAIGLALCCWEHIQPLSKYMYAD	180		
QY	181	EQVHVASPSPSLRGMAYALGPVNTAASQIYAVGGCYLASCATYSPKIKLVDP	240		
DB	181	EQVHVASPSPSLRGMAYALGPVNTAASQIYAVGGCYLASCATYSPKIKLVDP	240		
QY	241	DKEMPLKXAGGPFAMIFGPDGRLAERLPETBEGLLVADIDIGMIALAAADPAHYR	300		
DB	241	DKEMPLKXAGGPFAMIFGPDGRLAERLPETBEGLLVADIDIGMIALAAADPAHYR	300		

```
Db 241 DKEMFLKAGCGGAMTFCGPDGRALAEPLPTEBGLVADIDLCIMLAKRAADPAGHYSRP 300
Qy 301 DVTLLDRLRRPAQRVVTLDAAFEFQNEKDGAPALRVVAESAAAQ 346
Db 301 DVTLLDRLRRPAQRVVTLDAAFEFQNEKDGAPALRVVAESAAAQ 346

RESULT 2
Q6RMR4_9ZZZZ PRELIMINARY; PRT; 341 AA.
ID Q6RMR4_9ZZZZ
AC Q6RMR4;
DT 05-JUL-2004 (Tremblrel. 27, Created)
DT 05-JUL-2004 (Tremblrel. 27, Last sequence update)
DT 05-JUL-2004 (Tremblrel. 27, Last annotation update)
DE Nitrilase (EC 3.5.5.7).
GN ORFNames=BD7266;
OS unclutered organism.
OC unclassified; environmental samples.
OX NCBI_TaxID=155900;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX PubMed=15066841; DOI=10.1128/AEM.70.4.2429-2436.2004;
RA Robertson D.E., Chaplin J.A., Desantis G., Podar M., Madden M.,
RA Chi E., Richardson T., Milan A., Miller M., Weiner D.P., Wong K.,
RA McQuaid J., Farwell B., Preston L.A., Tan X., Sneed M.A., Keller M.,
RA Machur E., Kretz P.L., Burk M.J., Short J.M.;
RT "Exploring nitrilase sequence space for enantioselective catalysis.";
RL Appl. Environ. Microbiol. 70:2429-2436(2004).
DR EMBL; AY487438; AAR9385.1; -; Genomic DNA.
DR GO; GO:0018762; F:aliphatic nitrilase activity; IEA.
DR GO; GO:0016810; F:hydrolase activity, acting on carbon-nitrog. . .; IEA.
DR InterPro; IPR000132; N/CN hydrolase.
DR InterPro; IPR003010; N:nitrogen compound metabolism; IEA.
DR Pfam; PF00795; CN hydrolase; 1.
DR PROSITE; PS50263; CN HYDROLASE; 1.
DR PROSITE; PS00921; NITRIL_CHT_2; 1.
DR Hydrolase.
SQ SEQUENCE 341 AA; 36918 MW; 23AC5387F4C1FDEE CRC64;

Query March 81.1%; Score 1464; DB 2; Length 341;
Best Local Similarity 81.7%; Pred. No. 76-104;
Matches 272; Conservative 28; Mismatches 33; Indels 0; Gaps 0;

Qy 1 MSEBPTKRGAAVQAPVFLDRLTVERKAIGLIEQAAKQDVRLIAFPETWIPGYPFWIML 60
Db 1 MLSPVTQYRAAAVQAPSFLLDRIVEKTIATIEQAAGQDVRLIAFPETWIPGYLWIML 60
Qy 61 GAPAWGRFVQRYFENSILVRGSKQWQALADARRHGMHVAVAGYSERAGGSLYMGQATFGP 120
Db 61 GSPAWGMFVQRYFENSILVRGSKQWQALADARRHGMHVAVAGYSERAGGSLYMGQATFGP 120
Qy 121 DGDILIAARRKLPKTHAERTVFGEGDGSHLAVHDTAIGRLGCCHEHIOPLSKYAMVYAAD 180
Db 121 EGBELIAARRKLPKTHAERTVFGEGDGSHLAVHDTAIGRLGCCHEHIOPLSKYAMVYAAD 180
Qy 181 EQVHVASWPSFSILYRGMAVYALGPEVNTAASQIYAVEGGCYVLASCATVSPKIMIKVLVDP 240
Db 181 EQVHVASWPSFSILYRGMAVYALGPEVNTAASQIYAVEGGCYVLASCATVSPKIMIKVLVDP 240
Qy 241 DKEMFLKAGCGGAMTFCGPDGRALAEPLPTEBGLVADIDLCIMLAKRAADPAGHYSRP 300
Db 241 DKEMFLKAGCGGAMTFCGPDGRALAEPLPTEBGLVADIDLCIMLAKRAADPAGHYSRP 300
Qy 301 DVTLLDRLRRPAQRVVTLDAAFEFQNEKDGAP 333
Db 301 DVTLLDRLRRPAQRVVTLDAAFEFQNEKDGAP 333

RESULT 3
Q6RWS2_9ZZZZ PRELIMINARY; PRT; 348 AA.
ID Q6RWS2_9ZZZZ
AC Q6RWS2;
DT 05-JUL-2004 (Tremblrel. 27, Created)
DT 05-JUL-2004 (Tremblrel. 27, Last sequence update)
DT 05-JUL-2004 (Tremblrel. 27, Last annotation update)
DE Nitrilase (EC 3.5.5.7).
GN ORFNames=BD7753;
OS unclutered organism.
OC unclassified; environmental samples.
```

```
AC Q6RWS2;
DT 05-JUL-2004 (Tremblrel. 27, Created)
DT 05-JUL-2004 (Tremblrel. 27, Last sequence update)
DT 05-JUL-2004 (Tremblrel. 27, Last annotation update)
DE Nitrilase (EC 3.5.5.7).
GN ORFNames=BD5338;
OS unclutered organism.
OC unclassified; environmental samples.
OX NCBI_TaxID=155900;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX PubMed=15066841; DOI=10.1128/AEM.70.4.2429-2436.2004;
RA Robertson D.E., Chaplin J.A., Desantis G., Podar M., Madden M.,
RA Chi E., Richardson T., Milan A., Miller M., Weiner D.P., Wong K.,
RA McQuaid J., Farwell B., Preston L.A., Tan X., Sneed M.A., Keller M.,
RA Machur E., Kretz P.L., Burk M.J., Short J.M.;
RT "Exploring nitrilase sequence space for enantioselective catalysis.";
RL Appl. Environ. Microbiol. 70:2429-2436(2004).
DR EMBL; AY487430; AAR93737.1; -; Genomic DNA.
DR GO; GO:0016810; F:hydrolase activity, acting on carbon-nitrog. . .; IEA.
DR GO; GO:0006807; P:nitrogen compound metabolism; IEA.
DR InterPro; IPR000132; N/CN hydrolase.
DR InterPro; IPR003010; N:nitrogen compound metabolism; IEA.
DR InterPro; IPR002332; PII_GlnB_UWP_5.
DR Pfam; PF00795; CN hydrolase; 1.
DR PROSITE; PS50263; CN HYDROLASE; 1.
DR PROSITE; PS00920; NITRIL_CHT_1; 1.
DR PROSITE; PS00921; NITRIL_CHT_2; UNKNOWN 1.
DR PROSITE; PS00496; PII_GlnB_UWP; UNKNOWN 1.
DR Hydrolase.
SQ SEQUENCE 348 AA; 38397 MW; 95F18795D75C01A5 CRC64;

Query March 66.9%; Score 1209; DB 2; Length 348;
Best Local Similarity 67.1%; Pred. No. 2.5e-84;
Matches 230; Conservative 42; Mismatches 69; Indels 2; Gaps 2;

Qy 4 PMTK-YRGAAVQAPVFLDRLTVERKAIGLIEQAAKQDVRLIAFPETWIPGYPFWIMLGA 62
Db 2 PTKQFVAAVQAPVFLDLGALSKGISLIEBAASGAKLIAFPETWIPGYPFWIMLWDS 61
Qy 63 PAKMGRFVQRYFENSILVRGSKQWQALADARRHGMHVAVAGYSERAGGSLYMGQATFGP 122
Db 62 PAKMGRFVQRYFENSILVRGSKQWQALADARRHGMHVAVAGYSERAGGSLYMGQATFGP 122
Qy 123 DLIAARRKLPKTHAERTVFGEGDGSHLAVHDTAIGRLGCCHEHIOPLSKYAMVYAAD 182
Db 122 EITTRKRLKPTHAERTVFGEGDGSHLAVHDTAIGRLGCCHEHIOPLSKYAMVYAAD 181
Qy 183 VHVASWPSFSILYRGMAVYALGPEVNTAASQIYAVEGGCYVLASCATVSPKIMIKVLVDPDK 242
Db 182 IHISWPSFSILYRGMAVYALGPEVNTAASQIYAVEGGCYVLASCATVSPKIMIKVLVDPDK 241
Qy 243 EMPLKAGGGMFMTGPGGRALAEPLPTEBGLVADIDLCIMLAKRAADPAGHYSRPDV 302
Db 242 EPLLEBGGGFMVITGPGGRALAEPLPTEBGLVADIDLCIMLAKRAADPAGHYSRPDV 301
Qy 303 TRLLDRLRRPAQRVVTLDAAFEFQNEKDGAPALRVVAESAAAQ 345
Db 302 TRLLDRLRRPAQRVVTLDAAFEFQNEKDGAPALRVVAESAAAQ 343

RESULT 4
Q6RWP9_9ZZZZ PRELIMINARY; PRT; 353 AA.
ID Q6RWP9_9ZZZZ
AC Q6RWP9;
DT 05-JUL-2004 (Tremblrel. 27, Created)
DT 05-JUL-2004 (Tremblrel. 27, Last sequence update)
DT 05-JUL-2004 (Tremblrel. 27, Last annotation update)
DE Nitrilase (EC 3.5.5.7).
GN ORFNames=BD7753;
OS unclutered organism.
OC unclassified; environmental samples.
```

OX NCBI\_TaxID=155900;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RX PubMed=15066841; DOI=10.1128/AEM.70.4.2429-2436.2004;  
 RA Robertson D.E., Chaplin J.A., Desantis G., Podar M., Madden M.,  
 RA Chi E., Richardson T., Milan A., Miller M., Weiner D.P., Wong K.,  
 RA McQuaid J., Farwell B., Preston L.A., Tan X., Sneed M.A., Keller M.,  
 RA Mathur E., Kretz P.L., Burk M.J., Short J.M.;  
 RT "Exploring nitrilase sequence space for enantioselective catalysis.";  
 RL Appl. Environ. Microbiol. 70:2429-2436(2004).  
 DR EMBL: AY487543; AAR97490.1; -; Genomic DNA.  
 DR GO: GO:0018762; F:aliphatic nitrilase activity; IEA.  
 DR GO: GO:0016810; F:hydrolyase activity, acting on carbon-nitrog. .; IEA.  
 DR GO: GO:0006807; P:nitrogen compound metabolism; IEA.  
 DR InterPro: IPR000310; N/CN hydrolase.  
 DR InterPro: IPR000310; N/CN hydrolase.  
 DR Pfam: PF00795; CN hydrolase; 1.  
 DR PROSITE: PS50263; CN HYDROLASE; 1.  
 DR PROSITE: PS00920; NITRIL\_CHT\_1; 1.  
 DR PROSITE: PS00921; NITRIL\_CHT\_2; 1.  
 SQ SEQUENCE. 353 AA, 38029 MM; 997D13779FC31326 CRC64;  
 Query Match 64.5%; Score 1165; DB 2; Length 353;  
 Best Local Similarity 69.4%; Pred. No. 66-81;  
 Matches 229; Conservative 31; Mismatches 66; Indels 4; Gaps 3;  
 QY 7 KRGAAVQAAPVFLDLDRTEKAIIGLEQAAKQDVRLIAFPETWIPGYPMWLGAPAWG 66  
 DB 7 KKVAAVQAAPVFLDLDAVERAVERIDEAGAARLIAFPETWIPGYPMWLGAPAWA 66  
 QY 67 -MR-FVQRYFENSIVRGSKQWQALDAAARRHGMHVAGYSEPAAGSLYWGQALF 124  
 DB 67 IIRGFVSRFEDNSLYDSPQAEKLRRAAARRMNVVVLGSEPDGSLYIAQMIIGPDGT 126  
 QY 125 IAAARRKLTPTAERTVFGEGDGSHLAVHDTAIGRLGALCCWEHIQPLSKYAAADEQVH 184  
 DB 127 IAKRRKLTPTAERTVFGEGDGSHLAVHDTAIGRLGALCCWEHIQPLSKYAAADEQVH 186  
 QY 185 VASWPSFSIYRGMAVALGPEVNTAASQIYAVEGGCVLASCATVSPSEMIKVLVDPDKEM 244  
 DB 187 VAAWPSFSIYDPFPAHALGAEVNNAASKIYAVEGSCFVIAPCATVSGQAMIDELCTPEKIQ 246  
 QY 245 FLKAGGFPMTIRGPDGRALAEPLPTEREGLVADIDLGMIATAKAAADPAAGHSPDVT 304  
 DB 247 FLHAGGFPVITGPDGAPLAAPLPDKEGLYADIDLGMIATAKAAADPAAGHSPDVT 306  
 QY 305 LLLDRRPAORVVTLLDAAFEPQNEKGDAP 334  
 DB 307 LLFNNRPGIRVETM--ALPIDAETAAEAPA 334  
 RESULT 5  
 Q6RMFO\_92ZZZ PRELIMINARY; PRT; 353 AA.  
 ID Q6RMFO\_92ZZZ  
 AC Q6RMFO;  
 DT 05-JUL-2004 (TrEMBLrel. 27, Created)  
 DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)  
 DE 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)  
 DE Nitrilase (EC 3.5.5.7).  
 GN ORFNames=BD5217;  
 OS uncultured organism.  
 OC unclassified; environmental samples.  
 OX NCBI\_TaxID=155900;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RX PubMed=15066841; DOI=10.1128/AEM.70.4.2429-2436.2004;  
 RA Robertson D.E., Chaplin J.A., Desantis G., Podar M., Madden M.,  
 RA Chi E., Richardson T., Milan A., Miller M., Weiner D.P., Wong K.,  
 RA McQuaid J., Farwell B., Preston L.A., Tan X., Sneed M.A., Keller M.,  
 RA Mathur E., Kretz P.L., Burk M.J., Short J.M.;  
 RT "Exploring nitrilase sequence space for enantioselective catalysis.";  
 RL Appl. Environ. Microbiol. 70:2429-2436(2004).

DR EMBL: AY487543; AAR97499.1; -; Genomic DNA.  
 DR GO: GO:0018762; F:aliphatic nitrilase activity; IEA.  
 DR GO: GO:0016810; F:hydrolyase activity, acting on carbon-nitrog. .; IEA.  
 DR GO: GO:0006807; P:nitrogen compound metabolism; IEA.  
 DR InterPro: IPR000310; N/CN hydrolase.  
 DR InterPro: IPR000310; N/CN hydrolase.  
 DR Pfam: PF00795; CN hydrolase; 1.  
 DR PROSITE: PS50263; CN HYDROLASE; 1.  
 DR PROSITE: PS00920; NITRIL\_CHT\_1; 1.  
 DR PROSITE: PS00921; NITRIL\_CHT\_2; 1.  
 SQ SEQUENCE. 353 AA; 38321 MM; 2AA05257B3A7BC6C CRC64;  
 Query Match 64.0%; Score 1155; DB 2; Length 353;  
 Best Local Similarity 66.9%; Pred. No. 3.5e-80;  
 Matches 224; Conservative 38; Mismatches 69; Indels 4; Gaps 3;  
 QY 1 MSEPTKYRGAAPVFLDLDRTEKAIIGLEQAAKQDVRLIAFPETWIPGYPMWLG 60  
 DB 1 MSQIRPKLVAAVQAAPVFLDLDAVERAVERIDEAGAARLIAFPETWIPGYPMWLG 60  
 QY 61 GAPAWG-MR-FVQRYFENSIVRGSKQWQALDAAARRHGMHVAGYSEPAAGSLYWGQALF 118  
 DB 61 GAPAWA-MR-GFVSRYFEDNSLYDSPQAEKLRRAAARRMNVVVLGSEPDGSLYIAQMIIGPDGT 120  
 QY 119 GPDGDLIAARRKLTPTAERTVFGEGDGSHLAVHDTAIGRLGALCCWEHIQPLSKYAA 178  
 DB 121 GPGGRTVQRRKLTPTAERTVFGEGDGSHLAVHDTAIGRLGALCCWEHIQPLSKYAA 180  
 QY 179 ADEQVHVASWPSFSIYRGMAVALGPEVNTAASQIYAVEGGCVLASCATVSPSEMIKVLV 238  
 DB 181 QNEQIHVAWPSFSIYDPFPAHALGAEVNNAASKIYAVEGSCFVIAPCATVSGQAMIDELCD 240  
 QY 239 TPDKEMFLKAGGFPMTIRGPDGRALAEPLPTEREGLVADIDLGMIATAKAAADPAAGH 298  
 DB 241 TPEKHQFLHVGSGFVITGPDGSPPLAKPLPDKEGLYADIDLGMIATAKAAADPAAGH 300  
 QY 299 RPDVTIRLLDRRPAORVVTLLDAAFEPQNEKGDAP 333  
 DB 301 RPDVTIRLLFNNRPGIRVETM--ALPIDAETAAEAPA 333  
 RESULT 6  
 Q6RMFO\_92ZZZ PRELIMINARY; PRT; 358 AA.  
 ID Q6RMFO\_92ZZZ  
 AC Q6RMFO;  
 DT 05-JUL-2004 (TrEMBLrel. 27, Created)  
 DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)  
 DE 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)  
 DE Nitrilase (EC 3.5.5.7).  
 GN ORFNames=BD5263;  
 OS uncultured organism.  
 OC unclassified; environmental samples.  
 OX NCBI\_TaxID=155900;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RX PubMed=15066841; DOI=10.1128/AEM.70.4.2429-2436.2004;  
 RA Robertson D.E., Chaplin J.A., Desantis G., Podar M., Madden M.,  
 RA Chi E., Richardson T., Milan A., Miller M., Weiner D.P., Wong K.,  
 RA McQuaid J., Farwell B., Preston L.A., Tan X., Sneed M.A., Keller M.,  
 RA Mathur E., Kretz P.L., Burk M.J., Short J.M.;  
 RT "Exploring nitrilase sequence space for enantioselective catalysis.";  
 RL Appl. Environ. Microbiol. 70:2429-2436(2004).  
 DR EMBL: AY487543; AAR97499.1; -; Genomic DNA.  
 DR GO: GO:0018762; F:aliphatic nitrilase activity; IEA.  
 DR GO: GO:0016810; F:hydrolyase activity, acting on carbon-nitrog. .; IEA.  
 DR GO: GO:0006807; P:nitrogen compound metabolism; IEA.  
 DR InterPro: IPR000310; N/CN hydrolase.  
 DR InterPro: IPR000310; N/CN hydrolase.  
 DR Pfam: PF00795; CN hydrolase; 1.  
 DR PROSITE: PS50263; CN HYDROLASE; 1.  
 DR PROSITE: PS00920; NITRIL\_CHT\_1; 1.  
 DR PROSITE: PS00921; NITRIL\_CHT\_2; 1.

KW Hydrolyase. 358 AA; 38740 MW; A884DD406E6F44D CRC64;  
SQ SEQUENCE

Query Match 62.7%; Score 1132.5; DB 2; Length 358;  
Best Local Similarity 63.2%; Pred. No. 1.9e-78;  
Matches 220; Conservative 40; Mismatches 67; Indels 21; Gaps 5;

QY 7 KYRGAAYOAAPEVFLDRLTVEKAIGLIEQAAKQDVRLIAPPETWIPGYPFMIWLGAPAMG 66  
DB 7 KTKVAAYOAAPEFLDVAADVAVKAVLIDEAANGSSSLVAPEETWIPGYPFMIWLGAPAMA 66  
QY 67 -MR-FVQYFPENSLVKSGKQWQALADARRRGMHVAAGYSEBAGSLVMGQALRPPDDL 124  
DB 67 IMRGVSRYFPDNLSDYDQAEKRLDAKRRHKLTVWGLSEBAGSLYIAQMIIGPNET 126  
QY 125 IAAARKLKPTHAERTVFEEDGSHLAVHDTAIGRLGALCWEHIQPLSKYAMVAADQVH 184  
DB 127 VAQRKCLKPTHAERTVFEEDGSHLAVHNLPIGRGALCWEHIQPLSKYAMVAADQVH 186  
QY 185 VASWPSFSLYKGMAYALGPEVNTAASQIYAVEGSCYVLASCATVSPENIKVLDTPDKEM 244  
DB 187 VAAWPSFSLYDPFALHGAEVNNAASQIYAVEGSCFYVAPCAVISOEMIIDLCDTPDXHQ 246  
QY 245 FLKGGGPMIFGPDGALAEPLPETEEGLVADIDLGMIALAKAAADPAGHYSPDYTR 304  
DB 247 LIHVGSGTIVYIGPDGARGIKGLAPDQSGIVADIDLGMIPLAKAAADPAGHYARPDYTR 306  
QY 305 LLLDRRPAQRVYVTL-----DAFE---PQ-----NEKGDAP 333  
DB 307 LLENNRPAQRVETTLVPDQVARDIDARVEAALPQARPAATGNEPPAKP 354

RESULT 7  
Q6RMK6\_9ZZZZ PRELIMINARY; PRT; 354 AA.  
AC Q6RMK6;  
DT 05-JUL-2004 (TREMBLrel. 27, Created)  
DT 05-JUL-2004 (TREMBLrel. 27, Last sequence update)  
DT 05-JUL-2004 (TREMBLrel. 27, Last annotation update)  
DE Nitriylase (EC 3.5.5.7).  
GN ORFNames=BD5217;  
OS uncultured organism.  
OC unclassified; environmental samples.  
OX NCBI\_TaxID=155900;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RX PubMed=15066841; DOI=10.1128/JEM.70.4.2429-2436.2004;  
RA Robertson D.E., Chaplin J.A., Desantis G., Podar M., Madden M.,  
RA Chi E., Richardson T., Milan A., Miller M., Weiner D.P., Wong K.,  
RA McQuaid J., Kretz P.L., Preston L.A., Tan X., Sneed M.A., Keller M.,  
RA Machur E., Kretz P.L., Burk M.J., Short J.M.;  
RT "Exploring nitriylase sequence space for anantiselective catalysis.";  
RL Appl. Environ. Microbiol. 70:2429-2436 (2004).  
DR EMBL; AY487496; AAR97443.1; -; Genomic DNA.  
DR GO; GO:0018762; F:aliphatic nitriylase activity; IEA.  
DR GO; GO:0018810; F:hydrolyase activity, acting on carbon-nitrog. . .; IEA.  
DR GO; GO:0006807; P:nitrogen compound metabolism; IEA.  
DR InterPro; IPR000132; N/CN hydrolase.  
DR InterPro; IPR003010; Ntase/CNhydase.  
DR Pfam; PF00795; CN hydrolase; 1.  
DR PROSITE; PS50263; CN HYDROLASE; 1.  
DR PROSITE; PS00920; NITRIL\_CHT\_1; 1.  
DR PROSITE; PS00921; NITRIL\_CHT\_2; 1.  
KW Hydrolyase.  
SQ SEQUENCE 354 AA; 37879 MW; 8319C293775AE923 CRC64;

Query Match 62.6%; Score 1130; DB 2; Length 354;  
Best Local Similarity 64.4%; Pred. No. 2.9e-78;  
Matches 224; Conservative 42; Mismatches 74; Indels 8; Gaps 4;

QY 7 KYRGAAYOAAPEVFLDRLTVEKAIGLIEQAAKQDVRLIAPPETWIPGYPFMIWLGAPAMG 66  
DB 7 KTKVAAYOAAPEFLDVAADVAVKAVLIDEAANGSSSLVAPEETWIPGYPFMIWLGAPAMA 66

QY 67 -MR-FVQYFPENSLVKSGKQWQALADARRRGMHVAAGYSEBAGSLVMGQALRPPDDL 124  
DB 67 IMRGVSRYFPDNLSDYDQAEKRLDAKRRHKLTVWGLSEBAGSLYIAQMIIGPNET 126  
QY 125 IAAARKLKPTHAERTVFEEDGSHLAVHDTAIGRLGALCWEHIQPLSKYAMVAADQVH 184  
DB 127 VAQRKCLKPTHAERTVFEEDGSHLAVHNLPIGRGALCWEHIQPLSKYAMVAADQVH 186  
QY 185 VASWPSFSLYKGMAYALGPEVNTAASQIYAVEGSCYVLASCATVSPENIKVLDTPDKEM 244  
DB 187 VAAWPSFSLYDPFALHGAEVNNAASQIYAVEGSCFYVAPCAVISOEMIIDLCDTPDXHQ 246  
QY 245 FLKGGGPMIFGPDGALAEPLPETEEGLVADIDLGMIALAKAAADPAGHYSPDYTR 304  
DB 247 LIHVGSGTIVYIGPDGARGIKGLAPDQSGIVADIDLGMIPLAKAAADPAGHYARPDYTR 306  
QY 305 LLLDRRPAQRVYVTLDAFEPQNE--DKGDAPALRVAAESA-----AAQ 346  
DB 307 LLENNRPAQRVETTLALPVDQEAERAGAGKPAKSPVAALFTLTQAAE 354

RESULT 8  
Q706Q8\_PSEPU PRELIMINARY; PRT; 331 AA.  
AC Q706Q8;  
DT 05-JUL-2004 (TREMBLrel. 27, Created)  
DT 05-JUL-2004 (TREMBLrel. 27, Last sequence update)  
DT 05-JUL-2004 (TREMBLrel. 27, Last annotation update)  
DE Predicted amidohydrolyase/nitriylase.  
OC Pseudomonas putida.  
OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
OC Pseudomonadaceae; Pseudomonas.  
OX NCBI\_TaxID=303;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RX PubMed=9908981; PubMed=9791097;  
RA Ravatt R., Studer S., Zender A.J.B., van der Meer J.R.;  
RT "Int-B13, an unusual site-specific recombinase of the bacteriophage P4  
RT kb c1c-element of Pseudomonas sp. strain B13.";  
RL J. Bacteriol. 180:5505-5514 (1998).  
RN [2]  
RP NUCLEOTIDE SEQUENCE.  
RX PubMed=22708816; PubMed=12823813;  
RA Sentchilo V., Zehnder A.J.B., van der Meer J.R.;  
RT "Characterization of two alternative promoters and a transcription  
RT regulator for integrase expression in the c1c catabolic genomic island  
RT of Pseudomonas sp. strain B13.";  
RL Mol. Microbiol. 49:93-104 (2003).  
RN [3]  
RP NUCLEOTIDE SEQUENCE.  
RX STRAIN-strain RR21.  
RA Gaillard M., Merlen C., Vallaeys T., Vorhölter F.J., Pühler A.,  
RA van der Meer J.R.;  
RT Submitted (DEC-2003) to the EMBL/GenBank/DBJ databases.  
DR EMBL; AJ617740; CAB92892.1; -; Genomic DNA.  
DR GO; GO:0016810; F:hydrolyase activity, acting on carbon-nitrog. . .; IEA.  
DR GO; GO:0006807; P:nitrogen compound metabolism; IEA.  
DR InterPro; IPR000132; N/CN hydrolase.  
DR InterPro; IPR003010; Ntase/CNhydase.  
DR Pfam; PF00795; CN hydrolase; 1.  
DR PROSITE; PS50263; CN HYDROLASE; 1.  
DR PROSITE; PS00920; NITRIL\_CHT\_1; 1.  
DR PROSITE; PS00921; NITRIL\_CHT\_2; UNKNOWN 1.  
KW Hydrolyase.  
SQ SEQUENCE 331 AA; 36554 MW; DEF59D03874D9187 CRC64;

Query Match 62.4%; Score 1127; DB 2; Length 331;  
Best Local Similarity 65.2%; Pred. No. 4.5e-78;





Db 248 LVHAGGAAVIFGPDGRSLADKIPETQEGLLIYADIDGALGVAKNAADPAGHYSRPDVT 307  
QY 305 LLLDRRPAORV----VTLDAFEP 324  
Db 308 LLFNNKPARRVEYFSLPVDVAVETP 331

RESULT 11  
Q6RMW9\_9222Z PRELIMINARY; PRT; 337 AA.  
ID Q6RMW9\_9222Z PRELIMINARY; PRT; 337 AA.  
AC Q6RMW9;  
RX PubMed=15066841; DOI=10.1128/AEM.70.4.2429-2436.2004;  
RA Robertson D.E., Chaplin J.A., Desantis G., Podar M., Madden M.,  
RA Chl E., Richardson T., Milan A., Miller M., Weiner D.P., Wong K.,  
RA McQuaid J., Farwell B., Preston L.A., Tan X., Sneed M.A., Keller M.,  
RA Mathur E., Kretz P.L., Burk M.J., Short J.M.,  
RT "Exploring nitrilase sequence space for enantioselective catalysis."  
RL Appl. Environ. Microbiol. 70:2429-2436(2004).  
DR EMBL; AY487463; AAR97410.1; -; Genomic DNA.  
DR GO; GO:0018762; F:aliphatic nitrilase activity; IEA.  
DR GO; GO:0016810; F:hydrolase activity, acting on carbon-nitrog. .; IEA.  
DR GO; GO:0006807; P:nitrogen compound metabolism; IEA.  
DR InterPro: IPR000132; N/CN hydrolase.  
DR Pfam: PF00795; CN hydrolase; 1.  
DR PROSITE; PS50263; CN HYDROLASE; 1.  
DR PROSITE; PS00920; NITRIL\_CHT\_1; 1.  
DR PROSITE; PS00921; NITRIL\_CHT\_2; UNKNOWN\_1.  
KW Hydrolase.  
SQ SEQUENCE 337 AA; 36627 MW; A46AD7BD4047C1AC CRC64;

Query Match 59.6%; Score 1076.5; DB 2; Length 337;  
Best Local Similarity 67.6%; Pred. No. 3.4e-74;  
Matches 211; Conservative 33; Mismatches 65; Indels 3; Gaps 2;

QY 7 KYRGAAVQAAPVFLDRLRTVEKALGLEQAKQDVRLIAPPEWIPGYPFWIMLGAPAW- 65  
Db 7 KYRGAAVQAAPVFLDRLRTVEKALGLEQAKQDVRLIAPPEWIPGYPFWIMLGAPAW- 65  
QY 66 -GMRFVQRYFENSLVRGSKQWQALADAAARRHGMHVAVYSEBAGSLYMGQALFGPDGL 124  
Db 67 IGRGVQRYFENSLVRGSKQWQALADAAARRHGMHVAVYSEBAGSLYMGQALFGPDGL 124  
QY 125 IAAARKLKPTAERTVFGEDGSHLAVHDTA-IGRLGALCCWEHIQPLSKYAMTAADQV 183  
Db 127 IAKRRKLRPTAERTVFGEDGSHLAVHDTA-IGRLGALCCWEHIQPLSKYAMTAADQV 186  
QY 184 HVASWPSFSLYRGMAVYALGPEVNTAASQIYAVEGGCVYLASCATVSPEMIKVLVDPDKE 243  
Db 187 HVGAMPSPSLYDPFPAALGMEVNNAAASKYAVVSGSCFLGPAVVSQAMIDELCDSPDKH 246  
QY 244 MFLKAGGFPAMIFGPDGRALAEPLPETEGLLVADIDIGMIALAKAADPAGHYSRPDVT 303  
Db 247 AFLHAGGGFAIYVPPDSSSLAEKLAAPQEGLLVADIDIGMIALAKAADPAGHYSRPDVT 306  
QY 304 RLLDRRPAORV 315  
Db 307 RLLNTRPAARV 318

AC Q6RMW6;  
DT 05-JUL-2004 (TREMBlrel. 27, Created)  
DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)  
DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)  
DE Nitrilase (EC 3.5.7).  
GN ORFNames=BD5279;  
OS unclassified; environment. samples.  
OC unclassified; environment. samples.  
OX NCBI\_TaxId=155900;  
RN [1]  
RN NUCLEOTIDE SEQUENCE.  
RX PubMed=15066841; DOI=10.1128/AEM.70.4.2429-2436.2004;  
RA Robertson D.E., Chaplin J.A., Desantis G., Podar M., Madden M.,  
RA Chl E., Richardson T., Milan A., Miller M., Weiner D.P., Wong K.,  
RA McQuaid J., Farwell B., Preston L.A., Tan X., Sneed M.A., Keller M.,  
RA Mathur E., Kretz P.L., Burk M.J., Short J.M.,  
RT "Exploring nitrilase sequence space for enantioselective catalysis."  
RL Appl. Environ. Microbiol. 70:2429-2436(2004).  
DR EMBL; AY487536; AAR97483.1; -; Genomic DNA.  
DR GO; GO:0018762; F:aliphatic nitrilase activity; IEA.  
DR GO; GO:0016810; F:hydrolase activity, acting on carbon-nitrog. .; IEA.  
DR GO; GO:0006807; P:nitrogen compound metabolism; IEA.  
DR InterPro: IPR000132; N/CN hydrolase.  
DR InterPro: IPR003010; N/CN hydrolase.  
DR Pfam: PF00795; CN hydrolase; 1.  
DR PROSITE; PS50263; CN HYDROLASE; 1.  
DR PROSITE; PS00920; NITRIL\_CHT\_1; 1.  
DR PROSITE; PS00921; NITRIL\_CHT\_2; 1.  
KW Hydrolase.  
SQ SEQUENCE 338 AA; 36483 MW; B9EBD7FC216498BD CRC64;

Query Match 59.4%; Score 1073.5; DB 2; Length 338;  
Best Local Similarity 68.3%; Pred. No. 5.8e-74;  
Matches 213; Conservative 30; Mismatches 66; Indels 3; Gaps 2;

QY 7 KYRGAAVQAAPVFLDRLRTVEKALGLEQAKQDVRLIAPPEWIPGYPFWIMLGAPAW- 65  
Db 7 KYRGAAVQAAPVFLDRLRTVEKALGLEQAKQDVRLIAPPEWIPGYPFWIMLGAPAW- 65  
QY 66 -GMRFVQRYFENSLVRGSKQWQALADAAARRHGMHVAVYSEBAGSLYMGQALFGPDGL 124  
Db 67 IGRGVQRYFENSLVRGSKQWQALADAAARRHGMHVAVYSEBAGSLYMGQALFGPDGL 126  
QY 125 IAAARKLKPTAERTVFGEDGSHLAVHDTA-IGRLGALCCWEHIQPLSKYAMTAADQV 183  
Db 127 IAKRRKLRPTAERTVFGEDGSHLAVHDTA-IGRLGALCCWEHIQPLSKYAMTAADQV 186  
QY 184 HVASWPSFSLYRGMAVYALGPEVNTAASQIYAVEGGCVYLASCATVSPEMIKVLVDPDKE 243  
Db 187 HVGAMPSPSLYDPFPAALGMEVNNAAASKYAVVSGSCFLGPAVVSQAMIDELCDSPDKH 246  
QY 244 MFLKAGGFPAMIFGPDGRALAEPLPETEGLLVADIDIGMIALAKAADPAGHYSRPDVT 303  
Db 247 AFLHAGGGFAIYVPPDSSSLAEKLAAPQEGLLVADIDIGMIALAKAADPAGHYSRPDVT 306  
QY 304 RLLDRRPAORV 315  
Db 307 RLLNTRPAARV 318

RESULT 13  
Q6RMW6\_9222Z PRELIMINARY; PRT; 336 AA.  
ID Q6RMW6\_9222Z PRELIMINARY; PRT; 336 AA.  
AC Q6RMW6;  
RX PubMed=15066841; DOI=10.1128/AEM.70.4.2429-2436.2004;  
RA Robertson D.E., Chaplin J.A., Desantis G., Podar M., Madden M.,  
RA Chl E., Richardson T., Milan A., Miller M., Weiner D.P., Wong K.,  
RA McQuaid J., Farwell B., Preston L.A., Tan X., Sneed M.A., Keller M.,  
RA Mathur E., Kretz P.L., Burk M.J., Short J.M.,  
RT "Exploring nitrilase sequence space for enantioselective catalysis."  
RL Appl. Environ. Microbiol. 70:2429-2436(2004).  
DR EMBL; AY487536; AAR97483.1; -; Genomic DNA.  
DR GO; GO:0018762; F:aliphatic nitrilase activity; IEA.  
DR GO; GO:0016810; F:hydrolase activity, acting on carbon-nitrog. .; IEA.  
DR GO; GO:0006807; P:nitrogen compound metabolism; IEA.  
DR InterPro: IPR000132; N/CN hydrolase.  
DR InterPro: IPR003010; N/CN hydrolase.  
DR Pfam: PF00795; CN hydrolase; 1.  
DR PROSITE; PS50263; CN HYDROLASE; 1.  
DR PROSITE; PS00920; NITRIL\_CHT\_1; 1.  
DR PROSITE; PS00921; NITRIL\_CHT\_2; 1.  
KW Hydrolase.  
SQ SEQUENCE 338 AA; 36483 MW; B9EBD7FC216498BD CRC64;

Query Match 59.4%; Score 1073.5; DB 2; Length 338;  
Best Local Similarity 68.3%; Pred. No. 5.8e-74;  
Matches 213; Conservative 30; Mismatches 66; Indels 3; Gaps 2;

QY 7 KYRGAAVQAAPVFLDRLRTVEKALGLEQAKQDVRLIAPPEWIPGYPFWIMLGAPAW- 65  
Db 7 KYRGAAVQAAPVFLDRLRTVEKALGLEQAKQDVRLIAPPEWIPGYPFWIMLGAPAW- 66  
QY 66 -GMRFVQRYFENSLVRGSKQWQALADAAARRHGMHVAVYSEBAGSLYMGQALFGPDGL 124  
Db 67 IGRGVQRYFENSLVRGSKQWQALADAAARRHGMHVAVYSEBAGSLYMGQALFGPDGL 126  
QY 125 IAAARKLKPTAERTVFGEDGSHLAVHDTA-IGRLGALCCWEHIQPLSKYAMTAADQV 183  
Db 127 IAKRRKLRPTAERTVFGEDGSHLAVHDTA-IGRLGALCCWEHIQPLSKYAMTAADQV 186  
QY 184 HVASWPSFSLYRGMAVYALGPEVNTAASQIYAVEGGCVYLASCATVSPEMIKVLVDPDKE 243  
Db 187 HVGAMPSPSLYDPFPAALGMEVNNAAASKYAVVSGSCFLGPAVVSQAMIDELCDSPDKH 246  
QY 244 MFLKAGGFPAMIFGPDGRALAEPLPETEGLLVADIDIGMIALAKAADPAGHYSRPDVT 303  
Db 247 AFLHAGGGFAIYVPPDSSSLAEKLAAPQEGLLVADIDIGMIALAKAADPAGHYSRPDVT 306  
QY 304 RLLDRRPAORV 315  
Db 307 RLLNTRPAARV 318

RP NUCLEOTIDE SEQUENCE.  
 RX PubMed:15066841; DOI=10.1128/AEM.70.4.2429-2436.2004;  
 RA Robertson D.E., Chaplin J.A., Desantis G., Podar M., Madden M.,  
 RA Chi E., Richardson T., Milan A., Miller M., Weiner D.P., Wong K.,  
 RA McQuaid J., Fawell B., Preston L.A., Tan X., Sneed M.A., Keller M.,  
 RA Mathur E., Kretz P.L., Burk M.J., Short J.M.;  
 RT "Exploring nitrilase sequence space for enantioselective catalysis";  
 RL Appl. Environ. Microbiol. 70:2429-2436(2004).  
 DR EMBL: AY487546; AAR97493.1; -; Genomic DNA.  
 DR GO: GO:0018762; F:aliphatic nitrilase activity; IEA.  
 DR GO: GO:0016810; F:hydrolase activity, acting on carbon-nitrog. .; IEA.  
 DR GO: GO:0006807; P:nitrogen compound metabolism; IEA.  
 DR InterPro: IPR000132; N/CN hydrolase.  
 DR Pfam: PF00795; CN hydrolase; 1.  
 DR PROSITE: PS50263; CN HYDROLASE; 1.  
 DR PROSITE: PS00920; NITRIL\_CHT\_1; 1.  
 KW Hydrolase.  
 SQ SEQUENCE 336 AA; 36221 MW; CC370C188F94CB72 CRC64;  
 Query Match 59.2%; Score 1070; DB 2; Length 336;  
 Best Local Similarity 60.2%; Pred. No. 1.1e-73;  
 Matches 200; Conservative 53; Mismatches 75; Indels 4; Gaps 2;  
 QY 1 MSEPMTRYGAAYVQAAVFLDLDRTEKAIIGLEQAAKQVRLIAPFETWIPGYPMIWL 60  
 DB 3 ISHP--KKAAYVQAGPAFLDDGVEBAVSLIGAAEGQLIAFPETWIPGYPMIWL 60  
 QY 61 GAPAWKMF--FVORYFENSIVRGSKQWQALADAAARRHGMVAVGYSEBAGSLYMGQAI 118  
 DB 61 GSPAMAMKGFVQRFEDNALRHGSPQARISGAALAEHKIMVSLGFAERDGGTLVIAQWLI 120  
 QY 119 GPDDGLIAARRKLPKTHAERTVFGEGDGSILAVHDTAIGRLGALCCWEHIQPLSKYAMYA 178  
 DB 121 GPDGDTISRKKLKPETHERTVFGEGDSDLSVHDTAIGRLGALCCWEHIQPLSKYAMYA 180  
 QY 179 ADEQVHVASWPSFSLYRGMAVYALGPEVNTAASQIYAVEGCGCYVLASCATVSPENIKVLVD 238  
 DB 181 QNEQIHIGAMPFSLSYQPFANALSPENVIAVSRYAVAGCGCFPLAPCTVSDAMITELCD 240  
 QY 239 TPDKEMFLKAGGPFAMIFGPDGRALAEPLPETEGGLVADIDGMIALAKAADPAHYS 298  
 DB 241 TPEKQGLIRAGGHAALIFGPDGSLITPTVADTYEGILVYAEHDLGVIISAKAADPAHYS 300  
 QY 299 RPDVTRLILDRRPAQRYVTTLDAAFEPONEDKG 330  
 DB 301 RPDVTRLILNQTPSKRYQNMVLPLETYTEPBG 332  
 RESULT 14  
 Q6RMF7.9222Z PRELIMINARY; PRT; 335 AA.  
 AC Q6RMF7;  
 DT 05-JUL-2004 (TREMBlrel. 27, Created)  
 DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)  
 DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)  
 DE Nitrilase (EC 3.5.5.7).  
 GN ORFNames=BD5326;  
 OS unclassified organism.  
 OC unclassified; environmental samples.  
 OX NCBI\_TaxID=155900;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RX PubMed:15066841; DOI=10.1128/AEM.70.4.2429-2436.2004;  
 RA Robertson D.E., Chaplin J.A., Desantis G., Podar M., Madden M.,  
 RA Chi E., Richardson T., Milan A., Miller M., Weiner D.P., Wong K.,  
 RA McQuaid J., Fawell B., Preston L.A., Tan X., Sneed M.A., Keller M.,  
 RA Mathur E., Kretz P.L., Burk M.J., Short J.M.;  
 RT "Exploring nitrilase sequence space for enantioselective catalysis";  
 RL Appl. Environ. Microbiol. 70:2429-2436(2004).  
 DR EMBL: AY487545; AAR97492.1; -; Genomic DNA.  
 DR GO: GO:0018762; F:aliphatic nitrilase activity; IEA.  
 DR GO: GO:0016810; F:hydrolase activity, acting on carbon-nitrog. .; IEA.

DR GO: GO:0006807; P:nitrogen compound metabolism; IEA.  
 DR InterPro: IPR000132; N/CN hydrolase.  
 DR InterPro: IPR000132; N/CN hydrolase.  
 DR Pfam: PF00795; CN hydrolase; 1.  
 DR PROSITE: PS50263; CN HYDROLASE; 1.  
 DR PROSITE: PS00920; NITRIL\_CHT\_1; 1.  
 DR PROSITE: PS00921; NITRIL\_CHT\_2; 1.  
 KW Hydrolase.  
 SQ SEQUENCE 335 AA; 36902 MW; COABIDF5F963ACGA CRC64;  
 Query Match 59.2%; Score 1068.5; DB 2; Length 335;  
 Best Local Similarity 62.4%; Pred. No. 1.4e-73;  
 Matches 209; Conservative 42; Mismatches 79; Indels 5; Gaps 2;  
 QY 1 MSEPMTRYGAAYVQAAVFLDLDRTEKAIIGLEQAAKQVRLIAPFETWIPGYPMIWL 60  
 DB 1 MSITHPKFAAYVQAAVFLDLDSVKKALNLIIDEAAMAGAKLIAPFETWIPGYPMIWL 60  
 QY 61 GAPAW--GMEFVORYFENSIVRGSKQWQALADAAARRHGMVAVGYSEBAGSLYMGQAI 118  
 DB 61 GSPAMALGRGFVQRFEDNALRHGSPQADRLREARRNSITVVLGSRDGSILYIAQWLI 120  
 QY 119 GPDDGLIAARRKLPKTHAERTVFGEGDGSILAVHDTAIGRLGALCCWEHIQPLSKYAMYA 178  
 DB 121 GPDGDTISRKKLKPETHERTVFGEGDSDLSVHDTAIGRLGALCCWEHIQPLSKYAMYA 180  
 QY 179 ADEQVHVASWPSFSLYRGMAVYALGPEVNTAASQIYAVEGCGCYVLASCATVSPENIKVLVD 238  
 DB 181 QNEQIHIGAMPFSLSYQPFANALSPENVIAVSRYAVAGCGCFPLAPCTVSDAMITELCD 240  
 QY 239 TPDKEMFLKAGGPFAMIFGPDGRALAEPLPETEGGLVADIDGMIALAKAADPAHYS 298  
 DB 241 TPDKFELTHAGGHAALIFGPDGSLITPTVADTYEGILVYAEHDLGVIISAKAADPAHYS 300  
 QY 299 RPDVTRLILDRRPAQRYVTTLDAAFEPONEDKGAP 333  
 DB 301 RPDVTRLVLFMKIPAKRIEHFNL--FLDEQAGBEP 332  
 RESULT 15  
 Q5YUM5.NOCFA PRELIMINARY; PRT; 336 AA.  
 AC Q5YUM5;  
 DT 25-OCT-2004 (TREMBlrel. 28, Created)  
 DT 25-OCT-2004 (TREMBlrel. 28, Last sequence update)  
 DT 25-OCT-2004 (TREMBlrel. 28, Last annotation update)  
 DE Putative nitrilase.  
 GN OrderedLocuNames=nfa32690;  
 OS Nocardia farcinica.  
 OC Bacteria; Actinobacteria; Actinobacteridae; Actinomycetales;  
 OC Corynebacterineae; Nocardiaaceae; Nocardia.  
 OX NCBI\_TaxID=37329;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=IFM 10152;  
 RX PubMed:15466710; DOI=10.1073/pnas.0406410101;  
 RA Ishikawa J., Yamashita A., Mikami Y., Hoshino Y., Kurita H., Hotta K.,  
 RA Shiba T., Hatori M.;  
 RT "The complete genomic sequence of Nocardia farcinica IFM 10152";  
 RL Proc. Natl. Acad. Sci. U.S.A. 101:14925-14930(2004).  
 DR EMBL: AP006618; BAD58116.1; -; Genomic DNA.  
 DR GO: GO:0016810; F:hydrolase activity, acting on carbon-nitrog. .; IEA.  
 DR GO: GO:0006807; P:nitrogen compound metabolism; IEA.  
 DR InterPro: IPR000132; N/CN hydrolase.  
 DR InterPro: IPR002332; PII Glnb UMP\_S.  
 DR Pfam: PF00795; CN hydrolase; 1.  
 DR PROSITE: PS50263; CN HYDROLASE; 1.  
 DR PROSITE: PS00920; NITRIL\_CHT\_1; 1.  
 DR PROSITE: PS00921; NITRIL\_CHT\_2; 1.  
 DR PROSITE: PS00496; PII\_Glnb\_UMP; UNKNOWN\_1.  
 KW Complete proteome.  
 SQ SEQUENCE 336 AA; 36091 MW; D24921911BD0C2BB CRC64;

Query Match 59.2%; Score 1068.5; DB 2; Length 336;  
 Best Local Similarity 61.0%; Pred. No. 1.4e-73;  
 Matches 202; Conservative 47; Mismatches 79; Indels 3; Gaps 2;

QY	8	YRGAAYOAAFYFLDRTVEKAIGLIBOAAKODVRLTAFPEWTWIGYPFWTWLGAPAWGM	67
DB	7	FRAAVQAAPFVWLDGATVDKCVALLIEBALDNGAALLAFPETFVPGYFWTWLMDSPAWGM	66
QY	68	RFVQRYFENSLVRGSKOMQALADAAARRGHVAVYSEBAGSLYMGQALFGPDGLIAA	127
DB	67	QFVARYFDNSIALDGPLFARLREARARSATVTVTGHSEBDGSLYMGQALIGADDEVILAA	126
QY	128	RRKTKPTAERTVFGEDGSHLAVHDTAIGRLGALCCWEHTOPLSKYAMVAADQOVHAS	187
DB	127	RRKTKPTHVERTVFGESDGNLTVDTELGRGLGALCCWEHLQPLTKYAMTSQHEQIHVAA	186
QY	188	WPSFSLYRGMAYALGPEVNTAASQIYAVEGGCYVLASCATVSPMIKVLVDPDKEMFLK	247
DB	187	WPSFSVYRGAAYALGPEVNTGARQYAVEGQCFVLSPCAVIDEAGVELFCDTPAKRELL	246
QY	248	AGGGFAMIFGPDGRALAEPLPETEGLLVADIDGMIALAKAADPPAGHYSRPDYTRL	307
DB	247	PGGGFAQIYGPDGRELGTALPETEGLVYADLEASAVAVAKSAADPVGHYSRPVLOLLW	306
QY	308	DRRPAQ--RVVTLDAAFEPONEDKGDAPALR	336
DB	307	DRRPSVVRQVALSVASPASADAE-PAYR	336

Search completed: April 27, 2006, 01:00:10  
 Job time : 156.523 secs

XX



DR N-PSDB; ADH36217.  
XX Monitoring a chemical or biochemical process comprising providing a  
PT reactant comprising a cyanide or a material that can be converted to  
PT cyanide or a reactant that generates a cyanide or a material that can be  
PT converted to cyanide.  
XX  
PS Claim 73; SEQ ID NO 384; 277pp; English.  
XX  
CC This invention relates to a novel method of monitoring chemical or  
CC biochemical processes. The method involves providing a reactant  
CC comprising cyanide (or a material that can be converted to a cyanide)  
CC that generates as a reaction product cyanide or a material that can be  
CC converted to cyanide and measuring the concentration of produced cyanide.  
CC The method is useful for monitoring a chemical or biochemical process.  
CC The method is effective for high throughput systems and is sufficiently  
CC sensitive to detect a small amount of product. The present sequence is  
CC that of a nitrilase protein which may be used in the method of the  
CC invention.  
XX  
SQ Sequence 346 AA;  
Query Match 100.0%; Score 1806; DB 8; Length 346;  
Best Local Similarity 100.0%; Pred. No. 5,7e-172;  
Matches 346; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MSEPMTRYGAAYVQAAVFLDLDRTEKAIIGLEQAAKQDVRLIAPETWIPGYPMIWL 60  
Db 1 MSEPMTRYGAAYVQAAVFLDLDRTEKAIIGLEQAAKQDVRLIAPETWIPGYPMIWL 60  
QY 61 GAPAMGRRFQVRFFENSLVRSKKQWQALADARRHGMHVVAGYSERAGSLYMGQALFGP 120  
Db 61 GAPAMGRRFQVRFFENSLVRSKKQWQALADARRHGMHVVAGYSERAGSLYMGQALFGP 120  
QY 121 DGDILARRKLTPTHAERTVFGEGDGSILAVHDTAIGRLGALCWEHIQPLSKYAMTAAD 180  
Db 121 DGDILARRKLTPTHAERTVFGEGDGSILAVHDTAIGRLGALCWEHIQPLSKYAMTAAD 180  
QY 181 EGVHVASWPSFSLYRGMAVALGPEVNTAASQIYAVEGGCYVLASCATVSPMIKVLVDTP 240  
Db 181 EGVHVASWPSFSLYRGMAVALGPEVNTAASQIYAVEGGCYVLASCATVSPMIKVLVDTP 240  
QY 241 DKEMFLKAGGFMIFGPDGRALAEPLPETEEGLLVADIDLGMIATLAKAAADPAGHSR 300  
Db 241 DKEMFLKAGGFMIFGPDGRALAEPLPETEEGLLVADIDLGMIATLAKAAADPAGHSR 300  
QY 301 DVTRLIDRRRPAQRVVTLDAAFEPOEDKGDAPALRVVAESAAAQ 346  
Db 301 DVTRLIDRRRPAQRVVTLDAAFEPOEDKGDAPALRVVAESAAAQ 346  
RESULT 4  
ADG93917  
ID ADG93917 standard; protein; 346 AA.  
XX  
AC ADG93917;  
XX  
DT 11-MAR-2004 (first entry)  
XX  
DE Nitrilase enzyme amino acid sequence SegID384.  
XX  
KM nitrilase; nitrile; carboxylic acid; chemical process; pH; temperature;  
KM enantioselective transformation; enzyme.  
XX  
OS Unidentified.  
XX  
PN WO2003097810-A2.  
XX  
PD 27-NOV-2003.  
XX  
PF 15-MAY-2003; 2003WO-US015712.  
XX  
PR 15-MAY-2002; 2002US-00146772.

PR 09-SEP-2002; 2002US-00241742.  
XX  
PA (DIVE-) DIVERSA CORP.  
XX  
PI Desantis G, Short JM, Burk M, Wong K, Farwell R, Chatman K;  
XX  
DR WPI; 2004-090637/09.  
XX  
DR N-PSDB; ADG93916.  
XX  
PS Claim 44; SEQ ID NO 384; 295pp; English.  
XX  
CC This invention is related to a novel isolated or recombinant nucleic acid  
CC encoding a protein having nitrilase activity. Nitrilase s are capable of  
CC converting nitrile's directly to carboxylic acids and have great  
CC potential for use in industrial chemical processes. The isolated  
CC nitrilase proteins of the invention have increased activity and stability  
CC at increased pH and temperature when compared to those conventionally  
CC used. In addition, the nucleic acid of the invention is useful for  
CC screening enantioselective transformation. The present sequence is that  
CC of a nitrilase enzyme of the invention.  
XX  
SQ Sequence 346 AA;  
Query Match 100.0%; Score 1806; DB 8; Length 346;  
Best Local Similarity 100.0%; Pred. No. 5,7e-172;  
Matches 346; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MSEPMTRYGAAYVQAAVFLDLDRTEKAIIGLEQAAKQDVRLIAPETWIPGYPMIWL 60  
Db 1 MSEPMTRYGAAYVQAAVFLDLDRTEKAIIGLEQAAKQDVRLIAPETWIPGYPMIWL 60  
QY 61 GAPAMGRRFQVRFFENSLVRSKKQWQALADARRHGMHVVAGYSERAGSLYMGQALFGP 120  
Db 61 GAPAMGRRFQVRFFENSLVRSKKQWQALADARRHGMHVVAGYSERAGSLYMGQALFGP 120  
QY 121 DGDILARRKLTPTHAERTVFGEGDGSILAVHDTAIGRLGALCWEHIQPLSKYAMTAAD 180  
Db 121 DGDILARRKLTPTHAERTVFGEGDGSILAVHDTAIGRLGALCWEHIQPLSKYAMTAAD 180  
QY 181 EGVHVASWPSFSLYRGMAVALGPEVNTAASQIYAVEGGCYVLASCATVSPMIKVLVDTP 240  
Db 181 EGVHVASWPSFSLYRGMAVALGPEVNTAASQIYAVEGGCYVLASCATVSPMIKVLVDTP 240  
QY 241 DKEMFLKAGGFMIFGPDGRALAEPLPETEEGLLVADIDLGMIATLAKAAADPAGHSR 300  
Db 241 DKEMFLKAGGFMIFGPDGRALAEPLPETEEGLLVADIDLGMIATLAKAAADPAGHSR 300  
QY 301 DVTRLIDRRRPAQRVVTLDAAFEPOEDKGDAPALRVVAESAAAQ 346  
Db 301 DVTRLIDRRRPAQRVVTLDAAFEPOEDKGDAPALRVVAESAAAQ 346  
RESULT 5  
AD162516  
ID AD162516 standard; protein; 346 AA.  
XX  
AC AD162516;  
XX  
DT 22-APR-2004 (first entry)  
XX  
DE Nitrilase polypeptide #192.  
XX  
KM Atorvastatin; (R)-ethyl 4-cyano-3-hydroxybutyrate;  
KM (R)-ethyl 4-cyano-3-hydroxybutyric acid; epichlorohydrin;  
KM 3-hydroxyglutaronitrile; 3-hydroxyglutaronitrile;  
KM 4-cyano-3-hydroxybutyric acid; ethyl-4-cyano-3-hydroxybutyric acid;  
KM mixed hyperlipidaemia; homozygous familial hypercholesterolaemia;  
KM antihypertensive; enzyme.  
XX  
OS Unidentified.

```

XX  MO2003106415-A2.
XX
XX  24-DEC-2003.
XX
XX  13-JUN-2003; 2003WO-US018840.
XX
XX  13-JUN-2002; 2002US-0389317P.
XX
XX  28-JUN-2002; 2002US-0392944P.
XX
XX  (DIVE-) DIVERSA CORP.
XX
XX  Burk M, Desantis G, Morgan B, Zhu Z;
XX
XX  WPI; 2004-090821/09.
XX
XX  N-PSDB; ADI64637.
XX
XX  Preparation of atorvastatin comprises catalytic conversion of 3-
XX  hydroxyglutaronitrile by polypeptide with nitrilase activity, converting
XX  obtained 4-cyano-3-hydroxybutyric acid to ethyl-4-cyano-3-hydroxybutyric
XX  acid and forming atorvastatin.
XX
XX  Claim 41; SEQ ID NO 384; 253pp; English.
XX
XX  The present invention relates to a method for preparing an atorvastatin
XX  intermediate known as (R)-ethyl 4-cyano-3-hydroxybutyrate ((R)-ethyl 4-
XX  cyano-3-hydroxybutyric acid). The method comprises optionally converting
XX  epichlorohydrin or equivalent to 3-hydroxyglutaronitrile, catalytic
XX  conversion of 3-hydroxyglutaronitrile or equivalent to 4-cyano-3-
XX  hydroxybutyric acid with a polypeptide having nitrilase activity,
XX  converting 4-cyano-3-hydroxybutyric acid to ethyl-4-cyano-3-hydroxybutyric
XX  acid, and converting this to (R)-ethyl 4-cyano-3-hydroxybutyrate. The
XX  method involves whole cell processes, cell lysate processes, "one pot"
XX  processes, and "multi-pot" processes using a variety of parameters.
XX  Atorvastatin is used, in conjunction with dietary restriction, in the
XX  management of hyperlipidaemia, including hypercholesterolaemia, mixed
XX  dyslipidaemia and homozygous familial hypercholesterolaemia. The present
XX  sequence represents a nitrilase polypeptide obtained from an
XX  environmental sample.
XX
XX  Sequence 346 AA;
XX
XX  Query Match 100.0%; Score 1806; DB 8; Length 346;
XX  Best Local Similarity 100.0%; Pred. No. 5.7e-172;
XX  Matches 346; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX  1 MSEPMTKYRGAAVQAAVFLDLDRTEKAIIGLEQAKQDVRLIAFPETWIPGYPMIWL 60
XX  1 MSEPMTKYRGAAVQAAVFLDLDRTEKAIIGLEQAKQDVRLIAFPETWIPGYPMIWL 60
XX
XX  61 GAPAMGRPFQRYEENSLVSGSKOMQALDAARRHGMVAVGYSERAGSLVMGOAIFGP 120
XX  61 GAPAMGRPFQRYEENSLVSGSKOMQALDAARRHGMVAVGYSERAGSLVMGOAIFGP 120
XX
XX  121 DGDILAAARRLKPTHAERTVFGESGSHLAVHDTAIGRLGALCCWEHIQPLSKYAMYAAD 180
XX  121 DGDILAAARRLKPTHAERTVFGESGSHLAVHDTAIGRLGALCCWEHIQPLSKYAMYAAD 180
XX
XX  181 EOYVAVASWPSFSLYRGAAYALGPEVNTAAISOIYAVEGGCVVLASCAVSESEMIKVLVDP 240
XX  181 EOYVAVASWPSFSLYRGAAYALGPEVNTAAISOIYAVEGGCVVLASCAVSESEMIKVLVDP 240
XX
XX  241 DKEMFLKAGGAFAMIFGPDGRALAEPLPTEBEGLVADIDLGMIALAKAADPAHGYSRP 300
XX  241 DKEMFLKAGGAFAMIFGPDGRALAEPLPTEBEGLVADIDLGMIALAKAADPAHGYSRP 300
XX
XX  301 DVTIRLLIDRRPAORVVTLDAAFEQNDKDDAPRLRVVASSAAAO 346
XX  301 DVTIRLLIDRRPAORVVTLDAAFEQNDKDDAPRLRVVASSAAAO 346
XX
XX  RESULT 6
XX  ADI64637

```

```

ID  ADI64637 standard; protein; 346 AA.
XX
XX  ADI64637;
XX
XX  22-APR-2004 (first entry)
XX
XX  Nitrilase seq id 193.
XX
XX  (R)-ethyl 4-cyano-3-hydroxybutyric acid; nitrile hydrolysis;
XX  carboxylic acid; cyanohydrin moiety hydrolysis;
XX  ammonitrile moiety hydrolysis; chiral alpha-hydroxy acid molecule;
XX  chiral amino acid molecule; (S)-ethyl 4-cyano-3-hydroxybutyric acid;
XX  (R)-mandelic acid; (S)-mandelic acid; (S)-phenyl lactic acid derivative;
XX  (R)-phenyl lactic acid derivative; % enantiomeric excess; nitrilase.
XX
XX  Unidentified.
XX
XX  US2004014195-A1.
XX
XX  22-JUN-2004.
XX
XX  15-MAY-2003; 2003US-00440523.
XX
XX  29-DEC-1999; 99US-0173609P.
XX
XX  07-DEC-2000; 2000US-0254414P.
XX
XX  28-DEC-2000; 2000US-00751299.
XX
XX  21-JUN-2001; 2001US-0300189P.
XX
XX  30-JUL-2001; 2001US-0309006P.
XX
XX  22-JUN-2002; 2002US-0351336P.
XX
XX  15-MAY-2002; 2002US-00146772.
XX
XX  09-SEP-2002; 2002US-00241742.
XX
XX  (DIVE-) DIVERSA CORP.
XX
XX  Desantis G, Short JM, Burk MJ, Wong K, Farwell R, Chatman K;
XX
XX  WPI; 2004-121569/12.
XX
XX  N-PSDB; ADI64636.
XX
XX  Novel isolated or recombinant polypeptide having nitrilase activity,
XX  useful in production of food additives.
XX
XX  Claim 46; SEQ ID NO 384; 105pp; English.
XX
XX  The invention describes an isolated or recombinant polypeptide (I)
XX  comprising amino acids having a sequence at least 50 % identical to a
XX  sequence (SI) available in electronic form (EC) from the following web
XX  site ftp.segdata.uspto.gov/sequence.html?DocID=2004014195, or its
XX  variants, having one or more mutations at residue 55 Lys, Gly or Glu, at
XX  residue 60 glutamic acid, at residue 111 Ser, their combinations or
XX  fragments. (I) is useful for: producing an (R)-ethyl 4-cyano-3-
XX  hydroxybutyric acid; hydrolysing a nitrile to a carboxylic acid;
XX  hydrolysing cyanohydrin moiety or an ammonitrile moiety; producing a
XX  chiral alpha-hydroxy acid molecule or a chiral amino acid molecule;
XX  producing an (S)-ethyl 4-cyano-3-hydroxybutyric acid; producing an (R)-
XX  mandelic acid or (S)-mandelic acid; producing (S)-phenyl lactic acid
XX  derivative or an (R)-phenyl lactic acid derivative; modifying a molecule
XX  and for identifying a modified compound. The inventive method is useful
XX  for monitoring or determining % enantiomeric excess or % diastereomeric
XX  excess. (I) is useful in the production of food additives and drug
XX  intermediates. This is the amino acid sequence of a nitrilase of the
XX  invention.
XX
XX  Sequence 346 AA;
XX
XX  Query Match 100.0%; Score 1806; DB 8; Length 346;
XX  Best Local Similarity 100.0%; Pred. No. 5.7e-172;
XX  Matches 346; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
XX
XX  1 MSEPMTKYRGAAVQAAVFLDLDRTEKAIIGLEQAKQDVRLIAFPETWIPGYPMIWL 60
XX  1 MSEPMTKYRGAAVQAAVFLDLDRTEKAIIGLEQAKQDVRLIAFPETWIPGYPMIWL 60

```



QY 61 GAPAMGMRFFVQRYFENSIVRGSKOMQALADARRHGHWVAGYSEKAGSLYNGOAI FGP 120  
 DB 61 GAPAMGMRFFVQRYFENSIVRGSKOMQALADARRHGHWVAGYSEKAGSLYNGOAI FGP 120  
 QY 121 DGDLLIARRKLTKEPTHERTVFEGDGSHLAVHDTAIGRLGALCWEHIQPLSKYAMYAAD 180  
 DB 121 DGDLLIARRKLTKEPTHERTVFEGDGSHLAVHDTAIGRLGALCWEHIQPLSKYAMYAAD 180  
 QY 181 EQVHVASWPSFSLYRGMAVALGPEVNTAASQIYAVEGGCVVLASCATVSEPMIKVLVDT 240  
 DB 181 EQVHVASWPSFSLYRGMAVALGPEVNTAASQIYAVEGGCVVLASCATVSEPMIKVLVDT 240  
 QY 241 DKEMFLKAGGFMITFGPDGRALAEPLPETEEGLVADIDLGMITAKAADPA GHYSRP 300  
 DB 241 DKEMFLKAGGFMITFGPDGRALAEPLPETEEGLVADIDLGMITAKAADPA GHYSRP 300  
 QY 301 DVTRLLIDRRPAPQRYVTLDAAPFQNEKGDAPALRVAAESAAAQ 346  
 DB 301 DVTRLLIDRRPAPQRYVTLDAAPFQNEKGDAPALRVAAESAAAQ 346

RESULT 7  
 ADC23767  
 ID ADC23767 standard; protein; 341 AA.  
 AC ADC23767;  
 DT 18-DEC-2003 (first entry)  
 DE Protein sequence (SeqID 34) exhibiting nitrilase activity.  
 XX enzyme; nitrilase; nitrile; cyanohydrin; ammonia; biocatalyst;  
 KM enantiomer; chiral medicine.  
 XX Unidentified.  
 OS WO2003000840-A2.  
 PN 03-JAN-2003.  
 PD 15-MAY-2002; 2002WO-US015983.  
 PF 21-JUN-2001; 2001US-0300189P.  
 PR 30-JUL-2001; 2001US-0309006P.  
 PR 22-JAN-2002; 2002US-0351336P.  
 XX (DIVE-) DIVERSA CORP.  
 PA (MADD/) MADDEN D.  
 XX Madden M, Desantis G, Chaplin JA, Weiner DE, Milan A, Chi E;  
 PI Short JM, Burk M;  
 XX WPI; 2003-201417/19.  
 DR N-PSDB; ADC23766.  
 XX Novel nitrilase polypeptide, useful for making (R) - or (S) -ethyl-4-cyano-  
 PT 3-hydroxybutyric acid or (R) - or (S) -mandelic acid or (S) - or (R) -phenyl  
 PT lactic acid derivative and for producing pharmaceutical composition, and  
 PT food additive.  
 XX Claim 40; SEQ ID NO 34; 560pp; English.

CC inexpensive, efficient catalysts, have broad substrate specificity and  
 CC are capable of chiral differentiation. This polypeptide is a protein  
 CC sequence that exhibits nitrilase activity of the invention.  
 XX SQ Sequence 341 AA;  
 QY Query Match 81.1%; Score 1464; DB 7; Length 341;  
 DB Best Local Similarity 81.7%; Pred. No. 1,1e-137; Indels 0; Gaps 0;  
 DB Matches 272; Conservative 28; Mismatches 33; Indels 0; Gaps 0;  
 QY 1 MSEPTKRYGAAYQAAPVFLDIDRTVEKAIIGLIEQAAKQDVRLIAPETWPGYPFWIWL 60  
 DB 1 MLEPTQYRAAAYQAAPVFLDIDRTVEKTAIEQAABQDVRLIAPETWPGYPFWIWL 60  
 QY 61 GAPAMGMRFFVQRYFENSIVRGSKOMQALADARRHGHWVAGYSEKAGSLYNGOAI FGP 120  
 DB 61 GSPAMGMRFFVQRYFENSIVRGSKOMQALADARRHGHWVAGYSEKAGSLYNGOAI FGP 120  
 QY 121 DGDLLIARRKLTKEPTHERTVFEGDGSHLAVHDTAIGRLGALCWEHIQPLSKYAMYAAD 180  
 DB 121 EGELIARRKLTKEPTHERTVFEGDGSHLAVHDTAIGRLGALCWEHIQPLSKYAMYAAD 180  
 QY 181 EQVHVASWPSFSLYRGMAVALGPEVNTAASQIYAVEGGCVVLASCATVSEPMIKVLVDT 240  
 DB 181 EQVHVASWPSFSLYRGMAVALGPEVNTAASQIYAVEGGCVVLASCATVSEPMIKVLVDT 240  
 QY 241 DKEMFLKAGGFMITFGPDGRALAEPLPETEEGLVADIDLGMITAKAADPA GHYSRP 300  
 DB 241 DKEMFLKAGGFMITFGPDGRALAEPLPETEEGLVADIDLGMITAKAADPA GHYSRP 300  
 QY 301 DVTRLLIDRRPAPQRYVTLDAAPFQNEKGDAP 333  
 DB 301 DVTRLLIDRRPAPQRYVTLDAAPFQNEKGDAP 333

RESULT 8  
 ADH35868  
 ID ADH35868 standard; protein; 341 AA.  
 AC ADH35868;  
 DT 11-MAR-2004 (first entry)  
 DE Chemical process monitoring-related nitrilase protein sequence SeqID34.  
 XX Chemical process monitoring; biochemical process monitoring; cyanide;  
 KM high throughput system; enzyme.  
 XX Unidentified.  
 OS WO2003098187-A2.  
 PN 27-NOV-2003.  
 PD 15-MAY-2003; 2003WO-US015639.  
 PF 15-MAY-2002; 2002US-0380737P.  
 PR (DIVE-) DIVERSA CORP.  
 XX (DIVE-) DIVERSA CORP.  
 PA Weiner D, Chaplin JA, Chi E, Milan A, Desantis G, Burk M;  
 PI McQuaid J, Stege J;  
 XX WPI; 2004-142708/14.  
 DR N-PSDB; ADH35867.  
 XX Monitoring a chemical or biochemical process comprises providing a  
 PT reactant comprising a cyanide or a material that can be converted to  
 PT cyanide or a reactant that generates a cyanide or a material that can be  
 PT converted to cyanide.  
 XX Claim 73; SEQ ID NO 34; 277pp; English.



XX (DIVE-) DIVERSA CORP.  
 PA Burk M, Desantis G, Morgan B, Zhu Z;  
 PI WPI; 2004-090821/09.  
 DR N-PSDB; ADI62165.  
 XX  
 PT Preparation of atorvastatin comprises catalytic conversion of 3-  
 PT hydroxyglutaronitrile by polypeptide with nitrilase activity, converting  
 PT obtained 4-cyano-3-hydroxybutyric acid to ethyl-4-cyano-3-hydroxybutyric  
 PT acid and forming atorvastatin.  
 XX  
 PS Claim 41; SEQ ID NO 34; 253pp; English.  
 XX  
 CC The present invention relates to a method for preparing an atorvastatin  
 CC intermediate known as (R)-ethyl 4-cyano-3-hydroxybutyrate ((R)-ethyl 4-  
 CC cyano-3-hydroxybutyric acid). The method comprises optionally converting  
 CC epichlorohydrin or equivalent to 3-hydroxyglutaronitrile, catalytic  
 CC conversion of 3-hydroxyglutaronitrile or equivalent to 4-cyano-3-  
 CC hydroxybutyric acid with a polypeptide having nitrilase activity,  
 CC converting 4-cyano-3-hydroxybutyric acid to ethyl-4-cyano-3-hydroxybutyric  
 CC acid, and converting this to (R)-ethyl 4-cyano-3-hydroxybutyrate. The  
 CC method involves whole cell processes, cell lysate processes, "one pot"  
 CC processes, and "multi-pot" processes using a variety of parameters.  
 CC Atorvastatin is used, in conjunction with dietary restriction, in the  
 CC management of hyperlipidaemia, including hypercholesterolaemia, mixed  
 CC dyslipidaemia and homozygous familial hypercholesterolaemia. The present  
 CC sequence represents a nitrilase polypeptide obtained from an  
 CC environmental sample.  
 XX  
 SQ Sequence 341 AA;  
 XX  
 Query Match 81.1%; Score 1464; DB 8; Length 341;  
 Best Local Similarity 81.7%; Pred. No. 1.1e-137;  
 Matches 272; Conservative 28; Mismatches 33; Indels 0; Gaps 0;

KM carboxylic acid; cyanohydrin moiety hydrolysis;  
 KM aminonitrile moiety hydrolysis; chiral alpha-hydroxy acid molecule;  
 KM chiral amino acid molecule; (S)-ethyl 4-cyano-3-hydroxybutyric acid;  
 KM (R)-mandelic acid; (S)-mandelic acid; (S)-phenyl lactic acid derivative;  
 KM (R)-phenyl lactic acid derivative; \* enantiomeric excess;  
 KM \* diastereomeric excess; food additive; drug intermediate; nitrilase.  
 XX  
 OS Unidentified.  
 XX  
 XX US2004014195-A1.  
 EN  
 PD 22-JAN-2004.  
 XX  
 PF 15-MAY-2003; 2003US-00440523.  
 XX  
 PR 29-DEC-1999; 99US-0173609P.  
 PR 07-DEC-2000; 2000US-0254414P.  
 PR 28-DEC-2000; 2000US-00751299.  
 PR 21-JUN-2001; 2001US-0300189P.  
 PR 30-JUN-2001; 2001US-0309006P.  
 PR 22-JAN-2002; 2002US-0351336P.  
 PR 15-MAY-2002; 2002US-00146772.  
 PR 09-SEP-2002; 2002US-00241742.  
 XX  
 PA (DIVE-) DIVERSA CORP.  
 XX  
 PI Desantis G, Short JM, Burk MJ, Wong K, Farwell R, Chatman K;  
 XX WPI; 2004-121569/12.  
 DR N-PSDB; ADI64266.  
 XX  
 PT Novel isolated or recombinant polypeptide having nitrilase activity,  
 PT useful in production of food additives.  
 XX  
 PS Claim 46; SEQ ID NO 34; 105pp; English.  
 XX  
 CC The invention describes an isolated or recombinant polypeptide (I)  
 CC comprising amino acids having a sequence at least 50 % identical to a  
 CC sequence (S1) available in electronic form (EC) from the following web  
 CC site ftp.segdata.uspto.gov/sequence.html?docID=2004014195, or its  
 CC variants, having one or more mutations at residue 55 Iys, Gly or Glu, at  
 CC residue 60 glutamic acid, at residue 111 Ser, their combinations or  
 CC fragments. (I) is useful for: producing an (R)-ethyl 4-cyano-3-  
 CC hydroxybutyric acid; hydrolysing a nitrile to a carboxylic acid;  
 CC hydrolysing cyanohydrin moiety or an aminonitrile moiety; producing a  
 CC chiral alpha-hydroxy acid molecule or a chiral amino acid molecule;  
 CC producing an (S)-ethyl 4-cyano-3-hydroxybutyric acid; producing an (R)-  
 CC mandelic acid or (S)-mandelic acid; producing (S)-phenyl lactic acid  
 CC derivative or an (R)-phenyl lactic acid derivative; modifying a molecule;  
 CC and for identifying a modified compound. The inventive method is useful  
 CC for monitoring or determining \* enantiomeric excess or \* diastereomeric  
 CC excess. (I) is useful in the production of food additives and drug  
 CC intermediates. This is the amino acid sequence of a nitrilase of the  
 CC invention.  
 XX  
 SQ Sequence 341 AA;  
 XX  
 Query Match 81.1%; Score 1464; DB 8; Length 341;  
 Best Local Similarity 81.7%; Pred. No. 1.1e-137;  
 Matches 272; Conservative 28; Mismatches 33; Indels 0; Gaps 0;

```
QY 181 EQVHVASWPSFSLYRGMAYALGPEVNTAASQIYAVEGSCVYLASCATVSPEMIKVLVDP 240
DB 181 EQVHVASWPSFSLYRGMAYALGPEVNTAASQIYAVEGSCVYLASCATVSPEMIKVLVDP 240
QY 241 DDEMFLKAGGPFMIIRGPDGALAEPLPETEEGLVADIDLGMIATLAKKAADPAGHYSRP 300
DB 241 DDEMFLKAGGPFMIIRGPDGALAEPLPETEEGLVADIDLGMIATLAKKAADPAGHYSRP 300
QY 301 DVTLLDRLRRPAQRVVTLDAAFEFQNEDEKGDAP 333
DB 301 DVTLLDRLRRPAQRVVTLDAAFEFQNEDEKGDAP 333

RESULT 12
ADC23749
ID ADC23749 standard; protein; 348 AA.
XX
XX ADC23749;
XX
XX 18-DEC-2003 (first entry)
XX
DE Protein sequence (SeqID 16) exhibiting nitrilase activity.
XX
XX enzyme; nitrilase; nitrile; cyanohydrin; ammonia; biocatalyst;
XX enantiomer; chiral medicine.
XX
XX Unidentified.
XX
XX MO2003000840-A2.
XX
XX 03-JAN-2003.
XX
XX 15-MAY-2002; 2002WC-US015983.
XX
XX 21-JUN-2001; 2001US-0300189P.
XX 30-JUL-2001; 2001US-0309006P.
XX 22-JAN-2002; 2002US-0351336P.
XX
XX (DIVE-) DIVERSA CORP.
XX (MADD-) MADDEN D.
XX
XX Madden M, Desantis G, Chaplin JA, Weiner DF, Milan A, Chi E;
XX Short JM, Burk M;
XX
XX WPI: 2003-201417/19.
XX N-PSDB; ADC23748.
XX
XX Novel nitrilase polypeptide, useful for making (R) - or (S) -ethyl-4-cyano-
XX 3-hydroxybutyric acid or (R) - or (S) -mandelic acid or (S) - or (R) -phenyl
XX lactic acid derivative and for producing pharmaceutical composition, and
XX food additive.
XX
XX Claim 40; SEQ ID NO 16; 560pp; English.
XX
XX This invention relates to nitrilases and the nucleic acids that encode
XX these enzymes thereof. Specifically, it refers to polypeptides that
XX exhibit nitrilase activity, i.e. the ability to directly hydrolyse
XX nitriles or cyanohydrins into their corresponding carboxylic acids and
XX ammonia. Nitrilases have commercial utility as biocatalysts for use in
XX the synthesis of enantiomerically pure aromatic and aliphatic amino
XX acids, as well as hydroxy acids, which are important for the development
XX of chiral medicines. Furthermore, the present invention describes
XX nitrilases, isolated from mesophilic microorganisms, that have improved
XX activity and stability at increased pH and temperature. They are also
XX inexpensive, efficient catalysts, have broad substrate specificity and
XX are capable of chiral differentiation. This polypeptide is a protein
XX sequence that exhibits nitrilase activity of the invention.
XX
XX Sequence 348 AA;
XX
XX Query Match 66.9%; Score 1209; DB 7; Length 348;
XX Best Local Similarity 67.1%; Pred. No. 3.9e-112;
XX Matches 230; Conservative 42; Mismatches 69; Indels 2; Gaps 2;
```

```
QY 4 PMTK-YRGAAYAAVPELIDRTEKATIGLIEQAAKQDVRLIAFPEWTMIGYPFIMLGA 62
DB 2 PTKQFRVAAYAAVPELIDRTEKATIGLIEQAAKQDVRLIAFPEWTMIGYPFIMLGA 61
QY 63 PAMGRFVQRYFENSLVNGSKQWALADARRHGMVYVAGYERAGSLYMGQALFGPDG 122
DB 62 PAMGRFVQRYFENSLVNGSKQWALADARRHGMVYVAGYERAGSLYMGQALFGPDG 121
QY 123 DLIAARRKTKPTHARTVYGEDEGSHLVHDTALIRGLACGWEHIOPLSKXAMTAADPQ 182
DB 122 EFTTRRKLPHTHARTVYGEDEGSHLVHDTALIRGLACGWEHIOPLSKXAMTAADPQ 181
QY 183 VHVASWPSFSLYRGMAYALGPEVNTAASQIYAVEGSCVYLASCATVSPEMIKVLVDPDK 242
DB 182 IHASWPSFSLYRGAAYALGPELNNASQMYAAEGCFVLAPCATVSKEMIMLIDDPK 241
QY 243 EMFLKAGGPFMIIRGPDGALAEPLPETEEGLVADIDLGMIATLAKKAADPAGHYSRPV 302
DB 242 EPLLLEGGFMTIYGPDGRPLAKPLPENEEGLVADIDLGMIATLAKKAADPAGHYSRPV 301
QY 303 TRLLDRLRRPAQRVVTLDAAFEFQNEDEKGDAPALRYVASMAAA 345
DB 302 TRLLFMSAPANRVEYINPASGP-TESLKDMGKQWEAEQKAA 343

RESULT 13
ADH35850
ID ADH35850 standard; protein; 348 AA.
XX
XX ADH35850;
XX
XX 11-MAR-2004 (first entry)
XX
XX Chemical process monitoring-related nitrilase protein sequence SeqID16.
XX
XX chemical process monitoring; biochemical process monitoring; cyanide;
XX high throughput system; enzyme.
XX
XX Unidentified.
XX
XX WO2003098187-A2.
XX
XX 27-NOV-2003.
XX
XX 15-MAY-2003; 2003WC-US015639.
XX
XX 15-MAY-2002; 2002US-0380737P.
XX
XX (DIVE-) DIVERSA CORP.
XX
XX Weiner D, Chaplin JA, Chi E, Milan A, Desantis G, Burk MJ;
XX McQuaid J, Stege J;
XX
XX WPI: 2004-142708/14.
XX N-PSDB; ADH35849.
XX
XX Monitoring a chemical or biochemical process comprises providing a
XX reactant comprising a cyanide or a material that can be converted to
XX cyanide or a reactant that generates a cyanide or a material that can be
XX converted to cyanide.
XX
XX Claim 73; SEQ ID NO 16; 277pp; English.
XX
XX This invention relates to a novel method of monitoring chemical or
XX biochemical processes. The method involves providing a reactant
XX comprising cyanide (or a material that can be converted to a cyanide)
XX that generates as a reaction product cyanide or a material that can be
XX converted to cyanide and measuring the concentration of produced cyanide.
XX The method is useful for monitoring a chemical or biochemical process.
XX The method is effective for high throughput systems and is sufficiently
XX sensitive to detect a small amount of product. The present sequence is
XX that of a nitrilase protein which may be used in the method of the
```





GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

## OM protein - protein search, using sw model

Run on: April 27, 2006, 01:00:30 ; Search time 34.448 Seconds  
(without alignments)  
830.404 Million cell updates/sec

Title: US-09-751-299-2

Perfect score: 1806  
Sequence: 1 MSEPMTRYGAIVQAAPVFL.....EDKGDAPALRVVAESAAQAQ 346

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

## Database :

Issued Patents\_AA:\*  
1: /cgn2\_6/prodata/1/iaa/5\_COMB.pep:\*  
2: /cgn2\_6/prodata/1/iaa/6\_COMB.pep:\*  
3: /cgn2\_6/prodata/1/iaa/H\_COMB.pep:\*  
4: /cgn2\_6/prodata/1/iaa/PCTUS\_COMB.pep:\*  
5: /cgn2\_6/prodata/1/iaa/RE\_COMB.pep:\*  
6: /cgn2\_6/prodata/1/iaa/backfill1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	933	51.7	344	1	US-08-690-493-1 Sequence 1, Appl1
2	869	48.1	356	2	US-09-806-876A-2 Sequence 2, Appl1
3	807	44.7	369	2	US-09-823-373-5 Sequence 5, Appl1
4	807	44.7	369	2	US-09-823-373-14 Sequence 14, Appl1
5	592.5	32.8	354	1	US-08-447-702-5 Sequence 5, Appl1
6	592.5	32.8	354	1	US-08-465-615-5 Sequence 5, Appl1
7	176.5	9.8	325	2	US-09-543-681A-7440 Sequence 7440, Ap
8	150	8.3	315	2	US-10-105-294B-2 Sequence 2, Appl1
9	145.5	8.1	175	2	US-09-621-976-5160 Sequence 5160, Ap
10	144.5	8.0	337	2	US-09-489-039A-10443 Sequence 10443, Ap
11	135.5	7.5	302	2	US-09-252-991A-21992 Sequence 21992, A
12	131.5	7.3	152	2	US-09-621-976-7670 Sequence 7670, Ap
13	126	7.0	267	2	US-09-902-540-14601 Sequence 14601, A
14	124.5	6.9	501	2	US-09-863-339A-1 Sequence 1, Appl1
15	117.5	6.5	349	2	US-09-252-991A-24644 Sequence 24644, A
16	117	6.5	370	2	US-09-863-339A-2 Sequence 2, Appl1
17	116	6.4	664	2	US-09-902-540-12181 Sequence 12181, A
18	114.5	6.3	399	2	US-09-252-991A-28974 Sequence 28974, A
19	109	6.0	541	2	US-09-328-352-4728 Sequence 4728, Ap
20	108	6.0	513	2	US-09-949-016-6533 Sequence 6533, Ap
21	108	6.0	517	2	US-09-949-016-7160 Sequence 7160, Ap
22	104.5	5.8	409	1	US-08-809-740A-5 Sequence 5, Appl1
23	101.5	5.6	291	2	US-09-583-110-5130 Sequence 5130, Ap
24	101.5	5.6	294	2	US-09-107-433-4874 Sequence 4874, Ap
25	101.5	5.6	360	2	US-09-252-991A-22107 Sequence 22107, A
26	101	5.6	339	2	US-09-902-540-16772 Sequence 16772, A
27	101	5.6	633	2	US-09-252-991A-26229 Sequence 26229, A

28	100.5	5.6	540	2	US-09-252-991A-23300 Sequence 23300, A
29	100	5.5	648	2	US-09-902-540-14001 Sequence 14001, A
30	100	5.5	759	2	US-09-252-991A-133071 Sequence 133071, A
31	98	5.4	311	1	US-07-917-111-5 Sequence 5, Appl1
32	98	5.4	311	1	US-07-917-111-6 Sequence 6, Appl1
33	98	5.4	311	1	US-08-479-638-5 Sequence 5, Appl1
34	98	5.4	311	1	US-08-479-638-6 Sequence 6, Appl1
35	98	5.4	311	1	US-08-294-871A-70 Sequence 70, Appl1
36	98	5.4	311	2	US-08-876-398A-70 Sequence 70, Appl1
37	97.5	5.4	346	2	US-09-252-991A-27055 Sequence 27055, A
38	96	5.3	1990	2	US-09-502-540-11251 Sequence 11251, A
39	95	5.3	514	2	US-09-252-991A-22124 Sequence 22124, A
40	92.5	5.1	405	2	US-09-252-991A-23838 Sequence 23838, A
41	92.5	5.1	4563	2	US-09-108-006C-1 Sequence 1, Appl1
42	92	5.1	630	2	US-09-902-540-11585 Sequence 11585, A
43	92	5.1	666	2	US-09-252-991A-32525 Sequence 32525, A
44	91.5	5.1	401	2	US-09-252-991A-27198 Sequence 27198, A
45	91.5	5.1	415	2	US-08-861-774E-26 Sequence 26, Appl1

## ALIGNMENTS

RESULT 1  
US-08-690-493-1  
; Sequence 1, Application US/08690493  
; Patent No. 5872000  
GENERAL INFORMATION:  
; APPLICANT: Yu, Fujio  
; TITLE OF INVENTION: No. 5872000el Nitrlase Gene  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSER: Steinberg, Raakin & Davidson, P.C.  
; STREET: 1140 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 10036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: MS-DOS Editor  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/690,493  
; FILING DATE: 31 JUL-1996  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 213061/1995  
; FILING DATE: 31-JUL-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Davidson, Clifford M  
; REGISTRATION NUMBER: 32,728  
; REFERENCE/DOCKET NUMBER: 3821005  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 768-3800  
; TELEFAX: (212) 382-2124  
; INFORMATION FOR SEQ. ID NO. 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 344 amino acid residues  
; TYPE: amino acid  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; ORIGINAL SOURCE:  
; ORGANISM: Gordona terrae  
; STRAIN: MA-1  
; CELL TYPE: unicellular organism  
; FEATURE:  
; OTHER INFORMATION: Xaa is Met or a deletion  
; US-08-690-493-1  
Query Match 51.7%; Score 933; DB 1; Length 344;  
Best Local Similarity 55.2%; Pred. No. 7.5e+98;  
Matches 170; Conservative 52; Mismatches 86; Indels 0; Gaps 0;





```

:
: CURRENT APPLICATION NUMBER: US/09/823,373
:
: CURRENT FILING DATE: 2001-03-30
:
: PRIOR APPLICATION NUMBER: 60/193,707
:
: PRIOR FILING DATE: 2000-03-31
:
: NUMBER OF SEQ ID NOS: 32
:
: SOFTWARE: Microsoft Office 97
:
: SEQ ID NO 14
:
: LENGTH: 369
:
: TYPE: PRT
:
: ORGANISM: Acidovorax facilis
:
: US-09-823-373-14

```

Query Match	44.7%	Score 807	DB 2	length 369
Best Local Similarity	46.6%	Pred. No. 2.2e-83		
Matches 160	Conservative 52	Mismatches 127	Indels 4	Gaps 2

Qy	66	GMRPVORIFENSLVRGSKOMQALADARPHGMVAVGYSERAGSLYMGQALIPGGDGLI	125
Db	6	SKLTATVQAEPLWMDADATIDISIGIIEEAAOKGSLIAFPEYVITPCYPMANLGDVXY	65
Qy	6	TKYRGAAVQOAPVELDLDRIVEKALILIQAKOVRLLAPFETITPCYPMIWLGAAPW	65
Db	6	SKLTATVQAEPLWMDADATIDISIGIIEEAAOKGSLIAFPEYVITPCYPMANLGDVXY	65
Qy	66	GMRPVORIFENSLVRGSKOMQALADARPHGMVAVGYSERAGSLYMGQALIPGGDGLI	125
Db	66	SLSFSTRYHENSLELDDDRMRRLQTLAARNKIKALVMGYSERAGSRYLISQVIFDERGIBV	125
Qy	126	AASRRUKPTAERTVEGEDGSHLAVHPAIGRLALCWEHIQPLSKAMVYADDEVHV	185
Db	126	ANRRUKLPTHTVERTIYIGENGNDPLFHDPAFPRVGLGNWEHIFQLSKFMMYSIGDEVHV	185
Qy	186	ASMPSTSLRGMAIYALGPEVNTAASQIYAVBEGCYVLASCATVSPEMITKVLVTPDKEMF	245
Db	186	ASMPASMPLOPDVFOALISBANMTVTRSYALIEQOTFVLCTSTOYIGSALETFCCLNDEORAL	245
Qy	246	LKKGSGFAMIFCGEDGALAEPLPETEGGLVADIDILGMITALAKAADPAGYSRPDVTRL	305
Db	246	LPGCGCMARIYGPDSGEIANKPLAEDAEGLIYAEIDBGLILAKGADVGHYSRPDVLV	305
Qy	306	LILRR--PAQRYVTTLDAAPFEPONEDKGDAPALRYVASSAAAA	345
Db	306	QEPDRNHTTPVNR-IGIDGRILDVNTRSRYENFRLKQAAQEROA	347

RESULT 5  
 US-08-447-702-5  
 Sequence 5, Application US/08447702  
 Patent No. 5629190  
 GENERAL INFORMATION:  
 APPLICANT: Petre, Dominique  
 APPLICANT: Cerbeleaud, Edith  
 APPLICANT: Levy-Schill, Sophie  
 APPLICANT: Crouzet, Joel  
 TITLE OF INVENTION: Polypeptides Possessing A Nitrilase  
 TITLE OF INVENTION: Activity, DNA Sequence Coding for Said Polypeptides,  
 Patent No. 5629190  
 TITLE OF INVENTION: Expression Cassettes and Host Microorganisms Enabling Them  
 TITLE OF INVENTION: to be Obtained, and Method of Converting Nitriles to Carboxy  
 TITLE OF INVENTION: lates by Means of Said Polypeptides  
 NUMBER OF SEQUENCES: 5  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: BURNS, DOANE, SWECKER & MATTHIS  
 STREET: P O Box 1404  
 CITY: Alexandria  
 STATE: VA  
 COUNTRY: USA  
 ZIP: 22313-1404  
 COMPUTER READABLE FORM:  
 MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patentin Release #1.0, Version #1.30  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/447,702  
 FILING DATE: 23-MAY-1995  
 CLASSIFICATION: 530

```

1      PRIOR APPLICATION DATA:
2      APPLICATION NUMBER: US 08/194,588
3      FILING DATE: 10-FEB-1994
4      PRIOR APPLICATION DATA:
5      APPLICATION NUMBER: FR 92-09-882
6      FILING DATE: 10-AUG-1992
7      ATTORNEY/AGENT INFORMATION:
8      NAME: MCGOWAN, Malcolm K.
9      REGISTRATION NUMBER: 39,300
10     REFERENCE/DOCKET NUMBER: 003025-019
11     TELECOMMUNICATION INFORMATION:
12     TELEPHONE: 703-836-6620
13     TELEFAX: 703-836-2021
14     INFORMATION FOR SEQ ID NO: 5:
15     SEQUENCE CHARACTERISTICS:
16     LENGTH: 354 amino acids
17     TYPE: amino acid
18     TOPOLOGY: linear
19     MOLECULE TYPE: protein
20     US-08-447-702-5

```

Query Match	32.8%	Score 592.5	DB 1	Length 354
Best Local Similarity	39.4%	Pred. No. 7.2e-59		
Matches 137	Conservative 51	Mismatches 133	Indels 27	Gaps 7

```

OY 11 AAYQAAPEFLDLDRLVETKALIGIEBQAQODVRLIAPETWICGYPWIMLG-----APAW 65
Db 10 AAYQAAPEFNMLEAVTDCTCKLIAEAAANGAVIGFPEAFIFGYPWIMTISNDDFTGMMW 69
OY 66 GMRFVQVRFENSLVWGSKOMQALADAAARRHGHWVAGYAEERAGGSLYMGQAI RGPDDLI 125
Db 70 AV-----LEKALIEI PSKEVQOITSPAACKNGYVCVSEKONASLYTLQLMFDPGNLI 124
OY 126 AARRKLEKTHAERTVFGEGDGSHLVHDYTAGRLGALCEMEHIOPLSKYAMYAADREVHY 185
Db 125 GKHKREKTSSEBRAAWGWDGDSMAVFKTEYENLGGLOCEMEHALPNTIAAMSLNQVHY 184
OY 186 ASWSPF-----SLYRGMAVALGEVNTAASQIYAVEGCGYVLASCATVSPEMIKVLVD 238
Db 185 ASWPAFVFKGAVSSRVSSSVCSTAMHQIISOFAIISQVYVYVINSTNLVGGQIMIMIGK 244
OY 239 TPDKEMFLKAGGCFMTIFGPDGRALAEPLPETEBGLVADIDLGMALAKAAMDPRGHYS 298
Db 245 DEFSGNPLPLGSGNTAIISNTGEILIAS-IPOAEGIAVAEIDLNOIITYGKWLMDPAGHYS 303
OY 299 RPDVTRILLDR---PAQRV---VTLDAEAFQNEQDKG---APALRV 337
Db 304 TPFGSLSLTFDQSEHVPKKIGIBQTNHTSYEBLJHEDKDKMDLTIPRRV 351

```

RESULT 6  
 US-08-465-615-5  
 Sequence 5, Application US/08465615  
 Patent No. 5635391  
 GENERAL INFORMATION:  
 APPLICANT: PETRE, Dominique  
 APPLICANT: CERBELEAUD, Edith  
 APPLICANT: LEVY-SCHILL, Sophie  
 APPLICANT: CROUZET, Joel  
 TITLE OF INVENTION: POLYPEPTIDES POSSESSING A NITRILASE  
 TITLE OF INVENTION: ACTIVITY, DNA SEQUENCE CODING FOR SAID POLYPEPTIDES,  
 TITLE OF INVENTION: EXPRESSION CASSETTES AND HOST MICROORGANISMS ENABLING THEM  
 TITLE OF INVENTION: TO BE OBTAINED, AND METHOD OF CONVERTING NITRILES TO  
 TITLE OF INVENTION: CARBOXYLATES BY MEANS OF SAID POLYPEPTIDE  
 NUMBER OF SEQUENCES: 5  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Burns, Doane, Swecker & Mathis  
 STREET: P.O. Box 1404  
 CITY: Alexandria  
 STATE: Virginia  
 COUNTRY: United States  
 ZIP: 22313-1404  
 COMPUTER READABLE FORM:

```
/ MEDIUM TYPE: Floppy disk
/ COMPUTER: IBM PC compatible
/ OPERATING SYSTEM: PC-DOS/MS-DOS
/ SOFTWARE: Patentin Release #1.0, Version #1.25
/ CURRENT APPLICATION DATA:
/ APPLICATION NUMBER: US/08/465,615
/ FILING DATE:
/ CLASSIFICATION: 435
/ PRIORITY APPLICATION DATA:
/ APPLICATION NUMBER: US 08/194,588
/ FILING DATE: 10-FEB-1994
/ APPLICATION NUMBER: FR 9209882
/ FILING DATE: 10-AUG-1992
/ ATTORNEY/AGENT INFORMATION:
/ NAME: Crane-Feury, Sharon E
/ REGISTRATION NUMBER: 36,113
/ REFERENCE/DOCKET NUMBER: 003025-015
/ TELECOMMUNICATION INFORMATION:
/ TELEPHONE: (703) 836-6620
/ TELEFAX: (703) 836-2021
/ INFORMATION FOR SEQ ID NO: 5:
/ SEQUENCE CHARACTERISTICS:
/ LENGTH: 354 amino acids
/ TYPE: amino acid
/ TOPOLOGY: linear
/ MOLECULE TYPE: protein
/ US-08-465-615-5

Query Match      32.8%; Score 592.5; DB 1; Length 354;
Best Local Similarity 39.4%; Pred. No. 7.3e-59;
Matches 137; Conservative 51; Mismatches 133; Indels 27; Gaps 7;

QY 11 AAVQAAPVFLDRTVEKAIIGLEQAQKDVRLAEPETMIPGYPFWIWG-----APAW 65
DB 10 AAVQAAPVFLDRTVEKAIIGLEQAQKDVRLAEPETMIPGYPFWIWG-----APAW 69
QY 66 GKRFFVQRYFENSLVRGSKQWQALADARRHGMHVAVGYSERAGSLYMGQALFPGDGLI 125
DB 70 AV-----LFFNNAIEIPSRKVEQISDAKKNGVYVCVSVSKDNASLYLQGLWDPNGNII 124
QY 126 AARRKIKPHTAERTVFGEGDGSFLAVHDTAIGLGLCCWEHIQPSKAMVAAADQVRY 185
DB 125 GKRKRKPTFSERBAVVDGDSMAFVFKTEYGLGGLQCEHNLPLNTIAMGSLNQVHY 184
QY 186 ASWPSF-----SLYRGMAVALGPEVNTAASQIYAVEGGCYVLASCATVSPMIKVLVD 238
DB 185 ASWPAVVPKGAVSRRSSVSCASTNAMHQIISQFYAISNQVYIMSTNLVGDMDIMIGK 244
QY 239 TPDKEMFLKAGGGFAMIFGPDGRLAELPETEGLLVADIDIGMTALAKAAADPAQHY 298
DB 245 DESKKNFLPLGSGNTAIIISNTGELIAS-IPQDAEGIAVAEIDNLQIYCKMLDPAGHY 303
QY 299 RPDVTRLDDR---PAQRY---VTLDAPFEPQNEKGD---APALRY 337
DB 304 TEGFLSLTFDQSEHNVVKKIIGEQTNHFISTEDLHEDKMDMLTTPRRV 351

RESULT 7
US-09-543-681A-7440
/ Sequence 7440; Application US/09543681A
/ Patent No. 6605709
/ GENERAL INFORMATION:
/ APPLICANT: GARY BRETON
/ TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRABILIS
/ FILE REFERENCE: 2709.1002-001
/ CURRENT APPLICATION NUMBER: US/09/543,681A
/ PRIOR FILING DATE: 2000-04-05
/ PRIOR APPLICATION NUMBER: US 60/128,706
/ NUMBER OF SEQ ID NOS: 8344
/ SEQ ID NO 7440
/ LENGTH: 325
```

```
/ TYPE: PRT
/ ORGANISM: Proteus mirabilis
/ US-09-543-681A-7440

Query Match      9.8%; Score 176.5; DB 2; Length 325;
Best Local Similarity 24.7%; Pred. No. 2.6e-11;
Matches 85; Conservative 50; Mismatches 134; Indels 75; Gaps 15;

QY 4 PMTKRGAAPVQAAPVFLDRTVEKAIIGLEQAQKDVRLAEPETMIPGYPFWIWGAP 63
DB 5 PLTTLRVASVQLOHKANDKQVNLAKIHQETIEMASKEVNLVFEPCIRIGY--WHVPKLP 62
QY 64 AMCMRFVQRYFENSLVRGSKQWQALADARRHGMHVAVGYSER-AGGSILYMGQALFPGD 122
DB 63 A-----QQVYALSEKLAADSLSKSIKQKAQYAMALGVLEIRDNNNNLNTYTWVCMPPG 117
QY 123 DLIAARRKIKPHTA-ERTVFGEGDGSFLAVHDTAIGLGLCCWEHIQPSKAMVAA 178
DB 118 SL-----QKRRKLAHAFHPICGSD--QYTVFPTDPMQIKNGILLCDNDNLVENARATALLG 171
QY 179 AD-----EQVHVASPFSFLYRGMAVALGPEVNTAASQ----- 211
DB 172 ADILLAPHQGTGTHSRSPHMKRPIWALWEN-----RODDPOLQAAFQSEHGKWL 223
QY 212 -----IYAVEGGCYVLASCATVSPMIKVLVTPDKEMFLKAGGGFAMIFGPDGRLAEP 266
DB 224 KRWLPARAHNGWFFIFSG-----VGRDESEV---RTGNAMVIDPYGRIVKES 269
QY 267 LPETEGLLVADIDIGMTALAKAAADPAQHYSPDVTRLDDR 310
DB 270 C-AIEDMDVVTIDITLRLPESTGRRLTGR--RPELYQLTLTRQ 310

RESULT 8
US-10-105-294B-2
/ Sequence 2; Application US/10105294B
/ Patent No. 6800464
/ GENERAL INFORMATION:
/ APPLICANT: DRAUZ, KARLHEINZ
/ APPLICANT: MAY, OLIVER
/ APPLICANT: BOMMARTIUS, ANDREAS
/ APPLICANT: SYLDATK, CHRISTOPH
/ APPLICANT: ALTENBUCHNER, JOSEF
/ APPLICANT: WERNER, MARKUS
/ APPLICANT: SIEMANN-HERZBERG, MARTIN
/ TITLE OF INVENTION: D-Carbamoylase from Arthrobacter crystallopoietes
/ FILE REFERENCE: 220024USOX
/ CURRENT APPLICATION NUMBER: US/10/105,294B
/ PRIOR FILING DATE: 2000-08-11
/ PRIOR APPLICATION NUMBER: DE 101 14 999.9
/ NUMBER OF SEQ ID NOS: 14
/ SOFTWARE: Patentin version 3.1
/ SEQ ID NO 2
/ LENGTH: 315
/ TYPE: PRT
/ ORGANISM: ARTIFICIAL SEQUENCE
/ FEATURE:
/ OTHER INFORMATION: SYNTHETIC DNA
/ US-10-105-294B-2

Query Match      8.3%; Score 150; DB 2; Length 315;
Best Local Similarity 23.3%; Pred. No. 2.7e-08;
Matches 78; Conservative 42; Mismatches 107; Indels 108; Gaps 17;

QY 26 VEKATIGLEQAQKDVRLAEPETMIPGYPFWIWGAPAMGMPFVQRYFEN 76
DB 25 VAKTLALAEASQGAELVFPPELLTITFFPRTFEEGDF-----EEYFDK 70
QY 27 SLVRGSKQWQALADARRHGMHVAVGYSE-RAGGSILYMGQALFPGDGLIARRKIKL-PT 134
DB 71 SW--PNDVAPLPFERAKDGVGFYLGVALTSDKRYNSILVNKRGDILVGKRYRKHPLG 128
```



Db 233 -----LVFAGGSAAV--DPYQGLCE--AGRECEQIYBIDLG--RLQDARR 273  
Qy 292 D-----PAGHYSRDPVTRLIL 307  
Db 274 DYRYLEERLVLFGERRHDPDGLRELL 300

RESULT 12  
US-09-621-976-7670  
; Sequence 7670, Application US/09621976  
; Patent No. 6639063  
; GENERAL INFORMATION:  
; APPLICANT: Dumas Milne Edwards, J.B.  
; APPLICANT: Jobert, S.  
; APPLICANT: Giordano, J.Y.  
; TITLE OF INVENTION: ESTs and Encoded Human Proteins.  
; FILE REFERENCE: GENSET.054PR2  
; CURRENT APPLICATION NUMBER: US/09/621, 976  
; PRIOR FILING DATE: 2000-07-21  
; NUMBER OF SEQ ID NOS: 19335  
; SOFTWARE: Patent.pm  
; SEQ ID NO 7670  
; LENGTH: 152  
; TYPE: PRT  
; ORGANISM: Homo sapiens  
US-09-621-976-7670

Query Match 7.3%; Score 131.5; DB 2; Length 152;  
Best Local Similarity 27.8%; Pred. No. 1.1e-06;  
Matches 42; Conservative 24; Mismatches 60; Indels 25; Gaps 6;

Qy 5 MTKYRGAAVQAAVFLDLRTVEKAIGLIEQAAKQDVLRIAPETWIPGYPFWIMLGAPA 64  
Db 1 MTSFRLLALQLOISSIKSD-NVTRACSFIREATQAKIVSLPECNSPY----- 49

Qy 65 WGMRFQRYFENSIVGSKQWQALADARRHGMHVAG-YSEFAGSLYMGQAFGPDG 123  
Db 50 -GAKYFPEYAEK--IPG-ESTQKLSVAEKCSYILIGSIPBEDAKLYNTCAVFGPDGT 105

Qy 124 LIAARRKL-----KPTHAERTVFGSD 145  
Db 106 LIAKRYKIHLPDIDVPGKITPQESKTLSPGD 136

RESULT 13  
US-09-902-540-14601  
; Sequence 14601, Application US/09902540  
; Patent No. 6833447  
; GENERAL INFORMATION:  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Wiegand, Roger C.  
; TITLE OF INVENTION: Myxococcus xanthus Genome Sequences and Uses Thereof  
; FILE REFERENCE: 38-10(15849)B  
; CURRENT APPLICATION NUMBER: US/09/902, 540  
; PRIOR FILING DATE: 2001-07-10  
; PRIOR APPLICATION NUMBER: 60/217, 883  
; NUMBER OF SEQ ID NOS: 16825  
; SEQ ID NO 14601  
; LENGTH: 267  
; TYPE: PRT  
; ORGANISM: Myxococcus xanthus  
US-09-902-540-14601

Query Match 7.0%; Score 126; DB 2; Length 267;  
Best Local Similarity 23.8%; Pred. No. 1.1e-05;  
Matches 67; Conservative 35; Mismatches 110; Indels 70; Gaps 12;

Qy 21 DLDRIVEKAIGLIEQAAKQDVLRIAPETWIPGYPFWIMLGAPAWGMRFQRYFENSIVR 80  
Db 21 DLDRIVEKAIGLIEQAAKQDVLRIAPETWIPGYPFWIMLGAPAWGMRFQRYFENSIVR 80

Db 14 DKAHNLEAATRLVRAVVALGARLVGLPEN-----FSMMGPER-----ERODAAEGLD 60  
Qy 81 GSKQWQALADARRHGMHVAG--YSEFAGSLYMGQAFGPDGLIARRRLKPTHAE 137  
Db 61 GPTLSQ--MASLARLKVTLIAGSVLETGAPGGKLYNTSVLFGGGRILAVYRKI---HLF 116

Qy 138 RTVFGE-----DGSHLAVHDTAIGRLGALCWEHIOPLSKYANVADEQYHVAS 187  
Db 117 DVEVGDATYQSSAAVAPGTEVVAETEVGRGLSVCTDLRFP-ELYRLSREGATLLAV 175

Qy 188 WPSFSLYRGMAVALGPEVNTAASQIYAVEGCVYLAASCATVSEMIKVLVDPDKEMFLK 247  
Db 176 PAFTLMTGKHV---EVLRLAR--AIREQAVYLAFA----- 207

Qy 248 AGG-----GFAMIFGPDGRALAEPLPETEGGLVADID 280  
Db 208 QGGRHSANRVTYGHALVDPDWGLVLTAR--ASBEGIALAPVD 247

RESULT 14  
US-09-863-339A-1  
; Sequence 1, Application US/09863339A  
; Patent No. 6759524  
; GENERAL INFORMATION:  
; APPLICANT: Isobe, Minoru  
; APPLICANT: Matsuda, Tsukasa  
; TITLE OF INVENTION: Photoprotein Derived from Okinawan Squid and Gene  
; FILE REFERENCE: 026350-052  
; CURRENT APPLICATION NUMBER: US/09/863, 339A  
; PRIOR FILING DATE: 2001-05-24  
; PRIOR APPLICATION NUMBER: JP 2000-154786  
; PRIOR FILING DATE: 2000-05-25  
; PRIOR APPLICATION NUMBER: JP 2001-133941  
; NUMBER OF SEQ ID NOS: 13  
; SOFTWARE: Patentin version 3.1  
; SEQ ID NO 1  
; LENGTH: 501  
; TYPE: PRT  
; ORGANISM: Symplectoteuthis oualaniensis  
US-09-863-339A-1

Query Match 6.9%; Score 124.5; DB 2; Length 501;  
Best Local Similarity 18.7%; Pred. No. 4.6e-05;  
Matches 77; Conservative 57; Mismatches 110; Indels 167; Gaps 19;

Qy 4 PMTKYRGAAV--QAAVFLDLRTVEKAIGLIE-----QAAKQDVLRIAPETW 50  
Db 4 PVSWMKVAVFEHQVTPPKTDMETREBALDLKNSDVYHEAVLSRSKGVXKIVFPE-- 60

Qy 51 IGYPFWIMLGAPAWGMRFQRYFENSIVR-----G 81  
Db 61 -----YGL-----YDINTLTRTMDLMAEKVHPKIGHRNPCDEPEYQTS 101

Qy 82 SKQWQALADARRHGMHVAGYSER-----AGGS--LYMGQAFGPDGLIARR 128  
Db 102 SEMLRTPSCMAKENMVMVNNAGREPCRRATEPCGDKQLLYTNNVAFNNEGDAVARY 161

Qy 129 RLKLPKTHA--ERTVGEGBGSHLAVHDTAIGRLGALCWEHIOPLSKYANVADEQYHVA 186  
Db 162 YK---THLFWEBCWNRSSKNYEMALMDTPIGRFGTFMCPD-FQAVQLIEQYVNRHIAFPA 217

Qy 187 WPSF-----SLYRGMAVALGPEVNTAASQIYAVEGCVYLAASCATVSEMIKVLVDPD 241  
Db 218 SWNVLPPIYQISQISASAFARFAKINULAAASVHRLETSY----- 256

Qy 242 KEMFLKAGGFMATIGPDGRALAEPLPETEGGLVADIDLGKIALAKAAADPAGHSRPD 301  
Db 257 -----GSG---IYSPNG-----AEI-----FYFRD 274

Qy 302 V--TRLILD-----RRPAGRVTLLDAFEPQNE-----DKGDPALR 336  
Db 302 V--TRLILD-----RRPAGRVTLLDAFEPQNE-----DKGDPALR 336

Db 275 IPKSLIVAEILPIHVKKPEQTVVKNPNVPFSEDDVQDLFDRGDPFAFLK 325

RESULT 15  
US-09-252-991A-24644  
; Sequence 24644, Application US/09252991A  
; Patent No. 6551795  
; GENERAL INFORMATION:  
; APPLICANT: MARC J. RUBENFIELD et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; PRIOR FILING DATE: 1999-02-18  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 24644  
; LENGTH: 349  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-24644

Query Match 6.5%; Score 117.5; DB 2; Length 349;  
Best Local Similarity 25.8%; Pred. No. 0.00017;  
Matches 82; Conservative 22; Mismatches 99; Indels 115; Gaps 16;

QY 21 DLDRTVEKALIGLISOAKQDVRLIAFPETWIGYFPMWILGAPAMGRFVQRYFENSLVR 80  
Db 80 DVTANLAAARLLBOAEGKRLAVLPENF-----AAMGRDLAE-IGRAEAR 126  
QY 81 GS---KQWQALADARRHGMVAVGYSERAGSLYMGQALFGPDGLIA--ARRKUKPT 134  
Db 127 GNGPILPW--LNSAARDLRWIVA-----GTLPLPDGQPEAKANACSLIDE 172  
QY 135 HAERTV-----FGECD---GSHLAVHDTAIGRLGALCCWEHIQPLS 172  
Db 173 HGERVARYDKLHLPDVADARGRYRESDDYAFQKIVVADTPVGRIGLTVCYDLRFP-- 230  
QY 173 KYANYAADQGVHVASWPSFSLYRGMAVALGPEVNTAASQIYAVEGQ----- 218  
Db 231 -----ELYTMRRA-GAELITAPSATFAYTGAAMQVLVRAALET 270  
QY 219 -CYVLAS-CATVSPBEMIKVLVDTPDKEMFLKAGGFAMIFGPDGRALAEPLPETEGLLV 276  
Db 271 QCYLLAAGGGGVHPR-----GRETF-----GHSALVDPWGRVLAE-RPQGEAVILA 315  
QY 277 -----ADIDLGMITALA 287  
Db 316 VRDAAEQADIRRRMPVVA 333

Search completed: April 27, 2006, 01:02:21  
Job time : 35.448 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: April 27, 2006, 01:01:20 ; Search time 111.956 Seconds

(without alignments)  
1291.301 Million cell updates/sec

Title: US-09-751-299-2

Perfect score: 1806  
Sequence: 1 MSBPMYKRYRAAQAQAAVFL.....EDKGDAPALRVAAESAAQAQ 346

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :

Published Applications AA Main:  
1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBCOMB.pep:\*  
2: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBCOMB.pep:\*  
3: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBCOMB.pep:\*  
4: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBCOMB.pep:\*  
5: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBCOMB.pep:\*  
6: /cgn2\_6/ptodata/1/pubpaa/US11\_PUBCOMB.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1806	100.0	346	3	US-09-751-299-2
2	1806	100.0	346	4	US-10-146-772-384
3	1806	100.0	346	4	US-10-241-742-384
4	1806	100.0	346	4	US-10-440-523-384
5	1806	100.0	346	4	US-10-440-503-384
6	1806	100.0	346	4	US-10-461-925-384
7	1464	81.1	341	4	US-10-146-772-34
8	1464	81.1	341	4	US-10-241-742-34
9	1464	81.1	341	4	US-10-440-523-34
10	1464	81.1	341	4	US-10-440-503-34
11	1464	81.1	341	4	US-10-461-925-34
12	1209	66.9	348	4	US-10-146-772-16
13	1209	66.9	348	4	US-10-241-742-16
14	1209	66.9	348	4	US-10-440-523-16
15	1209	66.9	348	4	US-10-440-503-16
16	1209	66.9	348	4	US-10-461-925-16
17	1165	64.5	353	4	US-10-146-772-212
18	1165	64.5	353	4	US-10-241-742-212
19	1165	64.5	353	4	US-10-440-523-212
20	1165	64.5	353	4	US-10-440-503-212
21	1165	64.5	353	4	US-10-461-925-212
22	1155	64.0	353	4	US-10-146-772-232
23	1155	64.0	353	4	US-10-241-742-232
24	1155	64.0	353	4	US-10-440-523-232
25	1155	64.0	353	4	US-10-440-503-232
26	1155	64.0	353	4	US-10-461-925-232
27	1148	63.6	333	4	US-10-146-772-350

28	1148	63.6	333	4	US-10-241-742-350	Sequence 350, App
29	1148	63.6	333	4	US-10-440-523-350	Sequence 350, App
30	1148	63.6	333	4	US-10-461-925-350	Sequence 350, App
31	1148	63.6	333	4	US-10-146-772-170	Sequence 170, App
32	1132.5	62.7	358	4	US-10-241-742-170	Sequence 170, App
33	1132.5	62.7	358	4	US-10-440-523-170	Sequence 170, App
34	1132.5	62.7	358	4	US-10-461-925-170	Sequence 170, App
35	1132.5	62.7	358	4	US-10-146-772-170	Sequence 170, App
36	1132.5	62.7	358	4	US-10-440-523-170	Sequence 170, App
37	1130	62.6	354	4	US-10-146-772-102	Sequence 102, App
38	1130	62.6	354	4	US-10-241-742-102	Sequence 102, App
39	1130	62.6	354	4	US-10-440-523-102	Sequence 102, App
40	1130	62.6	354	4	US-10-461-925-102	Sequence 102, App
41	1130	62.6	354	4	US-10-146-772-334	Sequence 334, App
42	1125	62.3	345	4	US-10-241-742-334	Sequence 334, App
43	1125	62.3	345	4	US-10-440-523-334	Sequence 334, App
44	1125	62.3	345	4	US-10-461-925-334	Sequence 334, App
45	1125	62.3	345	4	US-10-146-772-334	Sequence 334, App

## ALIGNMENTS

```
RESULT 1
US-09-751-299-2
; Sequence 2, Application US/09751299
; Patent No. US2002001297A1
; GENERAL INFORMATION:
; APPLICANT: Maden, Mark
; APPLICANT: Weiner, David P.
; APPLICANT: Chaplin, Jennifer A.
; TITLE OR INVENTION: METHODS FOR PRODUCING ENANTIOMERICALLY PURE
; FILE REFERENCE: DIVER1440-2
; CURRENT APPLICATION NUMBER: US/09/751,299
; PRIOR APPLICATION NUMBER: 2000-12-29
; PRIOR FILING DATE: 2000-12-07
; PRIOR APPLICATION NUMBER: 60/173,609
; PRIOR FILING DATE: 1999-12-29
; NUMBER OF SEQ ID NOS: 4
; SOFTWARE: Patent In Ver. 2.1
; SEQ ID NO 2
; LENGTH: 346
; TYPE: PRT
; ORGANISM: Unknown Organism
; FEATURES:
; OTHER INFORMATION: Description of Unknown Organism: Obtained from an
; OTHER INFORMATION: environmental sample
US-09-751-299-2
Query Match 100.0%; Score 1806; DB 3; Length 346;
Best Local Similarity 100.0%; Pred. No. 2.4e-170;
Matches 346; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MSBPMYKRYRAAQAQAAVFLFDLDRTEKAIGLIEQAKQDVRLIAPETIIPGPFITWL 60
DB 1 MSBPMYKRYRAAQAQAAVFLFDLDRTEKAIGLIEQAKQDVRLIAPETIIPGPFITWL 60
QY 61 GAAWGRFQRFQFENSIVGSKOMQALDAAARRHGHVAVGYSBRAGSLYNGQALFGP 120
DB 61 GAAWGRFQRFQFENSIVGSKOMQALDAAARRHGHVAVGYSBRAGSLYNGQALFGP 120
QY 121 DGDILARRKLKPTAERTVFGEGDSHLAVHDTAIGRLGALCWEHIQPLSKYAMYAAD 180
DB 121 DGDILARRKLKPTAERTVFGEGDSHLAVHDTAIGRLGALCWEHIQPLSKYAMYAAD 180
QY 181 EQYHVASWPSFSLYRGMAVYALGPEVNTAASQIYAVBGGCYVLASCATVSPBMTKVLVDP 240
DB 181 EQYHVASWPSFSLYRGMAVYALGPEVNTAASQIYAVBGGCYVLASCATVSPBMTKVLVDP 240
QY 241 DKEMFLKAGGPFMTFGPDGRALAEPLPETEEGLVAVDIDLGMITALAKAANAPAGHSRP 300
```

Db 241 DKEMFLKAGGGFAMIFGPDGRALAEPLPETEEGLVADIDLGIMIALAKAAADPAGHSRP 300  
QY 301 DVTRLILDRRPAQRVVTLLDAAFEPQNEKGDAPALRVVAESAAAQ 346  
Db 301 DVTRLILDRRPAQRVVTLLDAAFEPQNEKGDAPALRVVAESAAAQ 346

RESULT 2  
US-10-146-772-384  
; Sequence 384, Application US/10146772  
; Publication No. US20030124698A1  
; GENERAL INFORMATION:  
; APPLICANT: Short, Jay  
; APPLICANT: Weiner, David  
; APPLICANT: Chaplin, Jennifer  
; APPLICANT: Chi, Ellen  
; APPLICANT: Milan, Aileen  
; APPLICANT: Desantis, Grace  
; APPLICANT: Madden, Mark  
; APPLICANT: Burk, Mark  
; TITLE OF INVENTION: Nitrlases  
; FILE REFERENCE: Docket No. US20030124698A1 DIV-013US  
; CURRENT FILING DATE: 2002-05-15  
; PRIOR FILING DATE: 2002-05-15  
; PRIOR APPLICATION NUMBER: US 60/309,006  
; PRIOR FILING DATE: 2001-07-30  
; PRIOR APPLICATION NUMBER: US 60/351,336  
; PRIOR FILING DATE: 2002-01-22  
; PRIOR APPLICATION NUMBER: US 60/300,189  
; PRIOR FILING DATE: 2001-06-21  
; PRIOR APPLICATION NUMBER: US 09/751,299  
; PRIOR FILING DATE: 2000-12-28  
; PRIOR APPLICATION NUMBER: US 60/254,414  
; PRIOR FILING DATE: 2000-12-07  
; PRIOR APPLICATION NUMBER: US 60/173,609  
; PRIOR FILING DATE: 1999-12-29  
; NUMBER OF SEQ ID NOS: 386  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 384  
; LENGTH: 346  
; TYPE: PRT  
; ORGANISM: Unknown  
; FEATURE:  
; OTHER INFORMATION: Obtained from an environmental sample  
US-10-146-772-384

Query Match 100.0%; Score 1806; DB 4; Length 346;  
Best Local Similarity 100.0%; Pred. No. 2.4e-170;  
Matches 346; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MSEPMTKYRGAAVOAAVFLDLDRIVEKAIGLIEQAQKQDVRLIAPETWIPGYPMIWL 60  
Db 1 MSEPMTKYRGAAVOAAVFLDLDRIVEKAIGLIEQAQKQDVRLIAPETWIPGYPMIWL 60  
QY 61 GAPAWGMRVQRYFENSIVRGSKQWQALADARRHGMHVAGYSEKAGSLTWGQAI FGP 120  
Db 61 GAPAWGMRVQRYFENSIVRGSKQWQALADARRHGMHVAGYSEKAGSLTWGQAI FGP 120  
QY 121 DGDLLIARRKLTPTAERTVFGEGDGSHTLVHDTAIGRLGALCWEHIQPLSKYAMYAAD 180  
Db 121 DGDLLIARRKLTPTAERTVFGEGDGSHTLVHDTAIGRLGALCWEHIQPLSKYAMYAAD 180  
QY 121 DGDLLIARRKLTPTAERTVFGEGDGSHTLVHDTAIGRLGALCWEHIQPLSKYAMYAAD 180  
Db 121 DGDLLIARRKLTPTAERTVFGEGDGSHTLVHDTAIGRLGALCWEHIQPLSKYAMYAAD 180  
QY 181 EQYHVASWPSFSLYRGMAYALGPEVNTAASQIYAAVEGGCYVLASCATVSPENIKVLVDP 240  
Db 181 EQYHVASWPSFSLYRGMAYALGPEVNTAASQIYAAVEGGCYVLASCATVSPENIKVLVDP 240  
QY 181 EQYHVASWPSFSLYRGMAYALGPEVNTAASQIYAAVEGGCYVLASCATVSPENIKVLVDP 240  
Db 181 EQYHVASWPSFSLYRGMAYALGPEVNTAASQIYAAVEGGCYVLASCATVSPENIKVLVDP 240  
QY 241 DKEMFLKAGGGFAMIFGPDGRALAEPLPETEEGLVADIDLGIMIALAKAAADPAGHSRP 300  
Db 241 DKEMFLKAGGGFAMIFGPDGRALAEPLPETEEGLVADIDLGIMIALAKAAADPAGHSRP 300  
QY 241 DKEMFLKAGGGFAMIFGPDGRALAEPLPETEEGLVADIDLGIMIALAKAAADPAGHSRP 300  
Db 241 DKEMFLKAGGGFAMIFGPDGRALAEPLPETEEGLVADIDLGIMIALAKAAADPAGHSRP 300  
QY 301 DVTRLILDRRPAQRVVTLLDAAFEPQNEKGDAPALRVVAESAAAQ 346  
Db 301 DVTRLILDRRPAQRVVTLLDAAFEPQNEKGDAPALRVVAESAAAQ 346

RESULT 3  
US-10-241-742-384  
; Sequence 384, Application US/10241742  
; Publication No. US20040002147A1  
; GENERAL INFORMATION:  
; APPLICANT: Short, Jay  
; APPLICANT: Weiner, David  
; APPLICANT: Chaplin, Jennifer  
; APPLICANT: Chi, Ellen  
; APPLICANT: Milan, Aileen  
; APPLICANT: Desantis, Grace  
; APPLICANT: Madden, Mark  
; APPLICANT: Burk, Mark  
; TITLE OF INVENTION: Nitrlases  
; FILE REFERENCE: Docket No. US20040002147A1 DIV-013US  
; CURRENT FILING DATE: 2002-09-09  
; PRIOR FILING DATE: 2002-05-15  
; PRIOR APPLICATION NUMBER: US/10/146,772  
; PRIOR FILING DATE: 2002-05-15  
; PRIOR APPLICATION NUMBER: US 60/309,006  
; PRIOR FILING DATE: 2001-07-30  
; PRIOR APPLICATION NUMBER: US 60/351,336  
; PRIOR FILING DATE: 2002-01-22  
; PRIOR APPLICATION NUMBER: US 60/300,189  
; PRIOR FILING DATE: 2001-06-21  
; PRIOR APPLICATION NUMBER: US 09/751,299  
; PRIOR FILING DATE: 2000-12-28  
; PRIOR APPLICATION NUMBER: US 60/254,414  
; PRIOR FILING DATE: 2000-12-07  
; PRIOR APPLICATION NUMBER: US 60/173,609  
; PRIOR FILING DATE: 1999-12-29  
; NUMBER OF SEQ ID NOS: 386  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 384  
; LENGTH: 346  
; TYPE: PRT  
; ORGANISM: Unknown  
; FEATURE:  
; OTHER INFORMATION: Obtained from an environmental sample  
US-10-241-742-384

Query Match 100.0%; Score 1806; DB 4; Length 346;  
Best Local Similarity 100.0%; Pred. No. 2.4e-170;  
Matches 346; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MSEPMTKYRGAAVOAAVFLDLDRIVEKAIGLIEQAQKQDVRLIAPETWIPGYPMIWL 60  
Db 1 MSEPMTKYRGAAVOAAVFLDLDRIVEKAIGLIEQAQKQDVRLIAPETWIPGYPMIWL 60  
QY 61 GAPAWGMRVQRYFENSIVRGSKQWQALADARRHGMHVAGYSEKAGSLTWGQAI FGP 120  
Db 61 GAPAWGMRVQRYFENSIVRGSKQWQALADARRHGMHVAGYSEKAGSLTWGQAI FGP 120  
QY 121 DGDLLIARRKLTPTAERTVFGEGDGSHTLVHDTAIGRLGALCWEHIQPLSKYAMYAAD 180  
Db 121 DGDLLIARRKLTPTAERTVFGEGDGSHTLVHDTAIGRLGALCWEHIQPLSKYAMYAAD 180  
QY 121 DGDLLIARRKLTPTAERTVFGEGDGSHTLVHDTAIGRLGALCWEHIQPLSKYAMYAAD 180  
Db 121 DGDLLIARRKLTPTAERTVFGEGDGSHTLVHDTAIGRLGALCWEHIQPLSKYAMYAAD 180  
QY 181 EQYHVASWPSFSLYRGMAYALGPEVNTAASQIYAAVEGGCYVLASCATVSPENIKVLVDP 240  
Db 181 EQYHVASWPSFSLYRGMAYALGPEVNTAASQIYAAVEGGCYVLASCATVSPENIKVLVDP 240  
QY 181 EQYHVASWPSFSLYRGMAYALGPEVNTAASQIYAAVEGGCYVLASCATVSPENIKVLVDP 240  
Db 181 EQYHVASWPSFSLYRGMAYALGPEVNTAASQIYAAVEGGCYVLASCATVSPENIKVLVDP 240  
QY 241 DKEMFLKAGGGFAMIFGPDGRALAEPLPETEEGLVADIDLGIMIALAKAAADPAGHSRP 300  
Db 241 DKEMFLKAGGGFAMIFGPDGRALAEPLPETEEGLVADIDLGIMIALAKAAADPAGHSRP 300  
QY 241 DKEMFLKAGGGFAMIFGPDGRALAEPLPETEEGLVADIDLGIMIALAKAAADPAGHSRP 300  
Db 241 DKEMFLKAGGGFAMIFGPDGRALAEPLPETEEGLVADIDLGIMIALAKAAADPAGHSRP 300  
QY 301 DVTRLILDRRPAQRVVTLLDAAFEPQNEKGDAPALRVVAESAAAQ 346  
Db 301 DVTRLILDRRPAQRVVTLLDAAFEPQNEKGDAPALRVVAESAAAQ 346

RESULT 4



```
US-10-440-523-384
; Sequence 384, Application US/10440523
; Publication No. US20040014195A1
; GENERAL INFORMATION:
; APPLICANT: Short, Jay
; APPLICANT: Weiner, David
; APPLICANT: Chaplin, Jennifer
; APPLICANT: Chi, Ellen
; APPLICANT: Milan, Aileen
; APPLICANT: Desantis, Grace
; APPLICANT: Madden, Mark
; APPLICANT: Burk, Mark
; TITLE OF INVENTION: Nitrites
; FILE REFERENCE: Docket No. US20040014195A1 DIV-013US
; CURRENT APPLICATION NUMBER: US/10/440,523
; CURRENT FILING DATE: 2003-05-15
; PRIOR APPLICATION NUMBER: US/10/146,772
; PRIOR FILING DATE: 2002-05-15
; PRIOR APPLICATION NUMBER: US 60/309,006
; PRIOR FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US 60/351,336
; PRIOR FILING DATE: 2002-01-22
; PRIOR APPLICATION NUMBER: US 60/300,189
; PRIOR FILING DATE: 2001-06-21
; PRIOR APPLICATION NUMBER: US 09/751,299
; PRIOR FILING DATE: 2000-12-28
; PRIOR APPLICATION NUMBER: US 60/254,414
; PRIOR FILING DATE: 2000-12-07
; PRIOR APPLICATION NUMBER: US 60/173,609
; PRIOR FILING DATE: 1999-12-29
; NUMBER OF SEQ ID NOS: 386
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 384
; LENGTH: 346
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Obtained from an environmental sample
US-10-440-523-384

Query Match      100.0%; Score 1806; DB 4; Length 346;
Best Local Similarity 100.0%; Pred. No. 2.4e-170;
Matches 346; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 MSEPMTKRGAAVOAPVFLDLDRVTEKAIGLEBOAKODVRLAEPETWIPGYFWIWL 60
        1 MSEPMTKRGAAVOAPVFLDLDRVTEKAIGLEBOAKODVRLAEPETWIPGYFWIWL 60
DB      1 MSEPMTKRGAAVOAPVFLDLDRVTEKAIGLEBOAKODVRLAEPETWIPGYFWIWL 60
QY      61 GAPAWGMRFVQRYFENSIVRSGSKOMQALADARRHGMHVAGYSERAGSLYMGQAIFGP 120
        61 GAPAWGMRFVQRYFENSIVRSGSKOMQALADARRHGMHVAGYSERAGSLYMGQAIFGP 120
DB      61 GAPAWGMRFVQRYFENSIVRSGSKOMQALADARRHGMHVAGYSERAGSLYMGQAIFGP 120
QY      121 DGDLLAARRKLPKPTAERTVFEGBGSHLAHVDTAIGLALCCWEHIQPLSKYAMYAAD 180
        121 DGDLLAARRKLPKPTAERTVFEGBGSHLAHVDTAIGLALCCWEHIQPLSKYAMYAAD 180
DB      121 DGDLLAARRKLPKPTAERTVFEGBGSHLAHVDTAIGLALCCWEHIQPLSKYAMYAAD 180
QY      181 EOVHVASPSSILYRGMAVYALGPDEVNTAASQIYAVEGGCYVLASCATVSPENIKYLVDP 240
        181 EOVHVASPSSILYRGMAVYALGPDEVNTAASQIYAVEGGCYVLASCATVSPENIKYLVDP 240
DB      181 EOVHVASPSSILYRGMAVYALGPDEVNTAASQIYAVEGGCYVLASCATVSPENIKYLVDP 240
QY      241 DKEMFLKAGGCFAMIFGPDGRALAEPLPETEGGLVADIDGMIALAKAADPAGHYSRP 300
        241 DKEMFLKAGGCFAMIFGPDGRALAEPLPETEGGLVADIDGMIALAKAADPAGHYSRP 300
DB      241 DKEMFLKAGGCFAMIFGPDGRALAEPLPETEGGLVADIDGMIALAKAADPAGHYSRP 300
QY      301 DVTRLILDRRPAQRVVTLDAAFEPOBEDKGPALRVVAESAAAQ 346
        301 DVTRLILDRRPAQRVVTLDAAFEPOBEDKGPALRVVAESAAAQ 346
DB      301 DVTRLILDRRPAQRVVTLDAAFEPOBEDKGPALRVVAESAAAQ 346

RESULT 5
US-10-440-503-384
; Sequence 384, Application US/10440503
; Publication No. US20040038419A1
```

```
; GENERAL INFORMATION:
; APPLICANT: Weiner, David Paul
; APPLICANT: Chaplin, Jennifer Ann
; APPLICANT: Chi, Ellen
; APPLICANT: Milan, Aileen
; APPLICANT: Desantis, Grace
; APPLICANT: Burk, Mark J.
; APPLICANT: McQuaid, Jeffrey
; APPLICANT: Stege, Uestlin
; TITLE OF INVENTION: ASSAYS AND KITS FOR DETECTING THE PRESENCE OF
; FILE REFERENCE: 09010-900001
; CURRENT APPLICATION NUMBER: US/10/440,503
; CURRENT FILING DATE: 2003-05-15
; PRIOR APPLICATION NUMBER: US 60/380,737
; PRIOR FILING DATE: 2002-05-15
; NUMBER OF SEQ ID NOS: 386
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 384
; LENGTH: 346
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Obtained from an environmental sample
US-10-440-503-384

Query Match      100.0%; Score 1806; DB 4; Length 346;
Best Local Similarity 100.0%; Pred. No. 2.4e-170;
Matches 346; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY      1 MSEPMTKRGAAVOAPVFLDLDRVTEKAIGLEBOAKODVRLAEPETWIPGYFWIWL 60
        1 MSEPMTKRGAAVOAPVFLDLDRVTEKAIGLEBOAKODVRLAEPETWIPGYFWIWL 60
DB      1 MSEPMTKRGAAVOAPVFLDLDRVTEKAIGLEBOAKODVRLAEPETWIPGYFWIWL 60
QY      61 GAPAWGMRFVQRYFENSIVRSGSKOMQALADARRHGMHVAGYSERAGSLYMGQAIFGP 120
        61 GAPAWGMRFVQRYFENSIVRSGSKOMQALADARRHGMHVAGYSERAGSLYMGQAIFGP 120
DB      61 GAPAWGMRFVQRYFENSIVRSGSKOMQALADARRHGMHVAGYSERAGSLYMGQAIFGP 120
QY      121 DGDLLAARRKLPKPTAERTVFEGBGSHLAHVDTAIGLALCCWEHIQPLSKYAMYAAD 180
        121 DGDLLAARRKLPKPTAERTVFEGBGSHLAHVDTAIGLALCCWEHIQPLSKYAMYAAD 180
DB      121 DGDLLAARRKLPKPTAERTVFEGBGSHLAHVDTAIGLALCCWEHIQPLSKYAMYAAD 180
QY      181 EOVHVASPSSILYRGMAVYALGPDEVNTAASQIYAVEGGCYVLASCATVSPENIKYLVDP 240
        181 EOVHVASPSSILYRGMAVYALGPDEVNTAASQIYAVEGGCYVLASCATVSPENIKYLVDP 240
DB      181 EOVHVASPSSILYRGMAVYALGPDEVNTAASQIYAVEGGCYVLASCATVSPENIKYLVDP 240
QY      241 DKEMFLKAGGCFAMIFGPDGRALAEPLPETEGGLVADIDGMIALAKAADPAGHYSRP 300
        241 DKEMFLKAGGCFAMIFGPDGRALAEPLPETEGGLVADIDGMIALAKAADPAGHYSRP 300
DB      241 DKEMFLKAGGCFAMIFGPDGRALAEPLPETEGGLVADIDGMIALAKAADPAGHYSRP 300
QY      301 DVTRLILDRRPAQRVVTLDAAFEPOBEDKGPALRVVAESAAAQ 346
        301 DVTRLILDRRPAQRVVTLDAAFEPOBEDKGPALRVVAESAAAQ 346
DB      301 DVTRLILDRRPAQRVVTLDAAFEPOBEDKGPALRVVAESAAAQ 346

RESULT 6
US-10-461-925-384
; Sequence 384, Application US/10461925
; Publication No. US20040053378A1
; GENERAL INFORMATION:
; APPLICANT: Mark J. Burk
; APPLICANT: Desantis, Grace
; APPLICANT: Morgan, Brian
; APPLICANT: Zhu, Zoulin
; TITLE OF INVENTION: PROCESSES FOR MAKING (R)-ETHYL 4-CYANO-3-HYDROXYBUTYRIC ACID
; FILE REFERENCE: 09010-270001
; CURRENT APPLICATION NUMBER: US/10/461,925
; CURRENT FILING DATE: 2003-06-13
; PRIOR APPLICATION NUMBER: US 60/389,317
; PRIOR FILING DATE: 2002-06-13
; PRIOR APPLICATION NUMBER: US 60/392,944
; PRIOR FILING DATE: 2002-06-28
; NUMBER OF SEQ ID NOS: 386
```

```
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 384
; LENGTH: 346
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Obtained from an environmental sample
US-10-461-925-384

Query Match      100.0%; Score 1806; DB 4; Length 346;
Best Local Similarity 100.0%; Pred. No. 2,4e-170;
Matches 346; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MSEPMTRYKGAAYOAAAPVFLDLDRTEKAIIGLEQAAKQDVRLIAPPETWIPGYPPIWTL 60
Db 1 MSEPMTRYKGAAYOAAAPVFLDLDRTEKAIIGLEQAAKQDVRLIAPPETWIPGYPPIWTL 60

QY 61 GAPAMGRFVQRYFENSLVRGSKOMQALADARRHGMHVAVGSEBAGGSLYMGQAIFGP 120
Db 61 GAPAMGRFVQRYFENSLVRGSKOMQALADARRHGMHVAVGSEBAGGSLYMGQAIFGP 120

QY 121 DGDILARRKLPKTHARTVFGEDGSHLVHDTAIGRLGALCWEHIQPLSKYAMYAAD 180
Db 121 DGDILARRKLPKTHARTVFGEDGSHLVHDTAIGRLGALCWEHIQPLSKYAMYAAD 180

QY 181 EQVHVASWPSFSLYRGMAVALGPEVNTAASQIYAVEGCVVLASCATVSPEMI KVLVDTP 240
Db 181 EQVHVASWPSFSLYRGMAVALGPEVNTAASQIYAVEGCVVLASCATVSPEMI KVLVDTP 240

QY 241 DKEMFLKAGGGFAMIFGPDGRALAEPLPETEBGLVADIDLGMTALAKAADPAHYSRP 300
Db 241 DKEMFLKAGGGFAMIFGPDGRALAEPLPETEBGLVADIDLGMTALAKAADPAHYSRP 300

QY 301 DVTRLILDRRPAORVVTLLDAAFEPQNEKGDAPALRVAAESAAAAQ 346
Db 301 DVTRLILDRRPAORVVTLLDAAFEPQNEKGDAPALRVAAESAAAAQ 346

RESULT 7
US-10-146-772-34
; Sequence 34, Application US/10146772
; Publication No. US20030124698A1
; GENERAL INFORMATION:
; APPLICANT: Short, Jay
; APPLICANT: Weiner, David
; APPLICANT: Chaplin, Jennifer
; APPLICANT: Chi, Ellen
; APPLICANT: Milani, Aileen
; APPLICANT: Desantis, Grace
; APPLICANT: Burk, Mark
; TITLE OF INVENTION: Nitrlases
; FILE REFERENCE: Docket No. US20030124698A1 DIV-013US
; CURRENT APPLICATION NUMBER: US/10/146,772
; CURRENT FILING DATE: 2002-05-15
; PRIOR APPLICATION NUMBER: US 60/309,006
; PRIOR FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US 60/351,336
; PRIOR FILING DATE: 2002-01-22
; PRIOR APPLICATION NUMBER: US 60/300,189
; PRIOR FILING DATE: 2001-06-21
; PRIOR APPLICATION NUMBER: US 09/751,299
; PRIOR FILING DATE: 2000-12-28
; PRIOR APPLICATION NUMBER: US 60/254,414
; PRIOR FILING DATE: 2000-12-07
; PRIOR APPLICATION NUMBER: US 60/173,609
; PRIOR FILING DATE: 1999-12-29
; NUMBER OF SEQ ID NOS: 386
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 34
; LENGTH: 341
; TYPE: PRT
; ORGANISM: Unknown
```

```
; FEATURE:
; OTHER INFORMATION: Obtained from an environmental sample
US-10-146-772-34

Query Match      81.1%; Score 1464; DB 4; Length 341;
Best Local Similarity 81.7%; Pred. No. 2.1e-136;
Matches 272; Conservative 28; Mismatches 33; Indels 0; Gaps 0;

QY 1 MSEPMTRYKGAAYOAAAPVFLDLDRTEKAIIGLEQAAKQDVRLIAPPETWIPGYPPIWTL 60
Db 1 MSEPMTRYKGAAYOAAAPVFLDLDRTEKAIIGLEQAAKQDVRLIAPPETWIPGYPPIWTL 60

QY 61 GAPAMGRFVQRYFENSLVRGSKOMQALADARRHGMHVAVGSEBAGGSLYMGQAIFGP 120
Db 61 GAPAMGRFVQRYFENSLVRGSKOMQALADARRHGMHVAVGSEBAGGSLYMGQAIFGP 120

QY 121 DGDILARRKLPKTHARTVFGEDGSHLVHDTAIGRLGALCWEHIQPLSKYAMYAAD 180
Db 121 DGDILARRKLPKTHARTVFGEDGSHLVHDTAIGRLGALCWEHIQPLSKYAMYAAD 180

QY 181 EQVHVASWPSFSLYRGMAVALGPEVNTAASQIYAVEGCVVLASCATVSPEMI KVLVDTP 240
Db 181 EQVHVASWPSFSLYRGMAVALGPEVNTAASQIYAVEGCVVLASCATVSPEMI KVLVDTP 240

QY 241 DKEMFLKAGGGFAMIFGPDGRALAEPLPETEBGLVADIDLGMTALAKAADPAHYSRP 300
Db 241 DKEMFLKAGGGFAMIFGPDGRALAEPLPETEBGLVADIDLGMTALAKAADPAHYSRP 300

QY 301 DVTRLILDRRPAORVVTLLDAAFEPQNEKGDAP 333
Db 301 DVTRLILDRRPAORVVTLLDAAFEPQNEKGDAP 333

RESULT 8
US-10-241-742-34
; Sequence 34, Application US/10241742
; Publication No. US20040002147A1
; GENERAL INFORMATION:
; APPLICANT: Short, Jay
; APPLICANT: Weiner, David
; APPLICANT: Chaplin, Jennifer
; APPLICANT: Chi, Ellen
; APPLICANT: Milani, Aileen
; APPLICANT: Desantis, Grace
; APPLICANT: Burk, Mark
; TITLE OF INVENTION: Nitrlases
; FILE REFERENCE: Docket No. US20040002147A1 DIV-013US
; CURRENT APPLICATION NUMBER: US/10/241,742
; CURRENT FILING DATE: 2002-09-09
; PRIOR APPLICATION NUMBER: US/10/146,772
; PRIOR FILING DATE: 2002-05-15
; PRIOR APPLICATION NUMBER: US 60/309,006
; PRIOR FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US 60/351,336
; PRIOR FILING DATE: 2002-01-22
; PRIOR APPLICATION NUMBER: US 60/300,189
; PRIOR FILING DATE: 2001-06-21
; PRIOR APPLICATION NUMBER: US 09/751,299
; PRIOR FILING DATE: 2000-12-28
; PRIOR APPLICATION NUMBER: US 60/254,414
; PRIOR FILING DATE: 2000-12-07
; PRIOR APPLICATION NUMBER: US 60/173,609
; PRIOR FILING DATE: 1999-12-29
; NUMBER OF SEQ ID NOS: 386
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 341
; LENGTH: 341
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Obtained from an environmental sample
US-10-241-742-34
```



```
QY 181 EOVHVASWPSFSLYRGMAVALGPEVNTAASQIYAVEGSCYVLASCATVSPEMIKVLVDT 240
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 181 EOVHVASWPSFSLYRGMAVALGPEVNTAASQVYAVEGSCYVLASCLVYTEILKVLIDTP 240
QY 241 DKEMFLKAGGPFAMI FGPDRALAEPLPETEEGLVADIDLGMLAKAADPA GHYSRP 300
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 241 DKEMFLKAGGPFAMI FGPDRALAEPLPETEEGLVADIDLGMLAKAADPA GHYAR 300
QY 301 DVTRLIDRRPAPQORVVTLDAAFEFQNEKGDAP 333
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 301 DVTRLIDRRPAPQORVVTLDAAFEFQNEKGDAP 333
```

```
RESULT 11
US-10-461-925-34
; Sequence 34, Application US/10461925
; Publication No. US20040053378A1
; GENERAL INFORMATION:
; APPLICANT: Mark J. Burk
; APPLICANT: Desantis, Grace
; APPLICANT: Morgan, Brian
; APPLICANT: Zhu, Zoulin
; TITLE OF INVENTION: PROCESSES FOR MAKING (R)-ETHYL 4-CYANO-3-HYDROXYBUTYRIC ACID
; FILE REFERENCE: 09010-270001
; CURRENT APPLICATION NUMBER: US/10/461,925
; PRIOR FILING DATE: 2003-06-13
; PRIOR APPLICATION NUMBER: US 60/389,317
; PRIOR FILING DATE: 2002-06-13
; PRIOR APPLICATION NUMBER: US 60/392,944
; PRIOR FILING DATE: 2002-06-28
; NUMBER OF SEQ ID NOS: 386
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 34
; LENGTH: 341
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Obtained from an environmental sample
US-10-461-925-34
```

```
Query Match 81.1%; Score 1464; DB 4; Length 341;
Best Local Similarity 81.7%; Pred. No. 2,1e-136;
Matches 272; Conservative 28; Mismatches 33; Indels 0; Gaps 0;

QY 1 MSEPMTYKRGAAVOAVALFVLDLDRTEVKAIGLIEQAAKQDRLIAFPETWIPGYPFMIWL 60
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 1 MSEPMTYKRGAAVOAVALFVLDLDRTEVKAIGLIEQAAKQDRLIAFPETWIPGYPFMIWL 60
QY 61 GAPAMGRFVQRYFENSLVSGSKQWQALADARRHGMHVAVGYSERAGGSLYMGQALFGP 120
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 61 GAPAMGRFVQRYFENSLVSGSKQWQALADARRHGMHVAVGYSERAGGSLYMGQALFGP 120
QY 121 DGLIAARRLKPTTHAERTVFGESDGSHLAVHDTAIGRLGALCCWEHIQPLSKYAMYAAD 180
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 121 DGLIAARRLKPTTHAERTVFGESDGSHLAVHDTAIGRLGALCCWEHIQPLSKYAMYAAD 180
QY 121 EGEILIAARRLKPTTHAERTVFGESDGSHLAVHDTAIGRLGALCCWEHIQPLSKYAMYAAD 180
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 121 EGEILIAARRLKPTTHAERTVFGESDGSHLAVHDTAIGRLGALCCWEHIQPLSKYAMYAAD 180
QY 181 EOVHVASWPSFSLYRGMAVALGPEVNTAASQIYAVEGSCYVLASCATVSPEMIKVLVDT 240
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 181 EOVHVASWPSFSLYRGMAVALGPEVNTAASQVYAVEGSCYVLASCLVYTEILKVLIDTP 240
QY 241 DKEMFLKAGGPFAMI FGPDRALAEPLPETEEGLVADIDLGMLAKAADPA GHYSRP 300
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 241 DKEMFLKAGGPFAMI FGPDRALAEPLPETEEGLVADIDLGMLAKAADPA GHYAR 300
QY 301 DVTRLIDRRPAPQORVVTLDAAFEFQNEKGDAP 333
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 301 DVTRLIDRRPAPQORVVTLDAAFEFQNEKGDAP 333
```

```
RESULT 12
US-10-146-772-16
; Sequence 16, Application US/10146772
; Publication No. US20030124698A1
```

```
; GENERAL INFORMATION:
; APPLICANT: Short, David
; APPLICANT: Weiner, David
; APPLICANT: Chaplin, Jennifer
; APPLICANT: Chi, Ellen
; APPLICANT: Milam, Aileen
; APPLICANT: Desantis, Grace
; APPLICANT: Madden, Mark
; APPLICANT: Burk, Mark
; TITLE OF INVENTION: Nitrlases
; FILE REFERENCE: Docket No. US20030124698A1 DIV-013US
; CURRENT APPLICATION NUMBER: US/10/146,772
; CURRENT FILING DATE: 2002-05-15
; PRIOR APPLICATION NUMBER: US 60/309,006
; PRIOR FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US 60/351,336
; PRIOR FILING DATE: 2002-01-22
; PRIOR APPLICATION NUMBER: US 60/300,189
; PRIOR FILING DATE: 2001-06-21
; PRIOR APPLICATION NUMBER: US 09/751,299
; PRIOR FILING DATE: 2000-12-28
; PRIOR APPLICATION NUMBER: US 60/254,414
; PRIOR FILING DATE: 2000-12-07
; PRIOR APPLICATION NUMBER: US 60/173,609
; PRIOR FILING DATE: 1999-12-29
; NUMBER OF SEQ ID NOS: 386
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 348
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Obtained from an environmental sample
US-10-146-772-16
```

```
Query Match 66.9%; Score 1209; DB 4; Length 348;
Best Local Similarity 67.1%; Pred. No. 4.5e-111;
Matches 230; Conservative 42; Mismatches 69; Indels 2; Gaps 2;

QY 4 PMTK-YRGAVALFVLDLDRTEVKAIGLIEQAAKQDRLIAFPETWIPGYPFMIWLGA 62
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 2 PMTK-YRGAVALFVLDLDRTEVKAIGLIEQAAKQDRLIAFPETWIPGYPFMIWLGA 62
QY 63 PAMGRFVQRYFENSLVSGSKQWQALADARRHGMHVAVGYSERAGGSLYMGQALFGP 122
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 63 PAMGRFVQRYFENSLVSGSKQWQALADARRHGMHVAVGYSERAGGSLYMGQALFGP 122
QY 123 DLIAARRLKPTTHAERTVFGESDGSHLAVHDTAIGRLGALCCWEHIQPLSKYAMYAAD 182
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 123 DLIAARRLKPTTHAERTVFGESDGSHLAVHDTAIGRLGALCCWEHIQPLSKYAMYAAD 182
QY 122 ETLFTTRKALKPTTHAERTVFGESDGSHLAVHDTAIGRLGALCCWEHIQPLSKYAMYAAD 181
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 122 ETLFTTRKALKPTTHAERTVFGESDGSHLAVHDTAIGRLGALCCWEHIQPLSKYAMYAAD 181
QY 183 VHVASWPSFSLYRGMAVALGPEVNTAASQIYAVEGSCYVLASCATVSPEMIKVLVDT 242
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 183 VHVASWPSFSLYRGMAVALGPEVNTAASQVYAVEGSCYVLASCLVYTEILKVLIDTP 242
QY 182 IHASWPSFSLYRGMAVALGPEVNTAASQVYAVEGSCYVLASCATVSPEMIKVLVDT 241
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 182 IHASWPSFSLYRGMAVALGPEVNTAASQVYAVEGSCYVLASCATVSPEMIKVLVDT 241
QY 243 EMFLKAGGPFAMI FGPDRALAEPLPETEEGLVADIDLGMLAKAADPA GHYSRP 302
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 243 EMFLKAGGPFAMI FGPDRALAEPLPETEEGLVADIDLGMLAKAADPA GHYAR 302
QY 242 EPLLEGGGFTMTIGDGRPLAKPLPENEGLVADIDLGMLAKAADPA GHYAR 301
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 242 EPLLEGGGFTMTIGDGRPLAKPLPENEGLVADIDLGMLAKAADPA GHYAR 301
QY 303 TRLLIDRRPAPQORVVTLDAAFEFQNEKGDAPALRVVAESAAAA 345
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 303 TRLLIDRRPAPQORVVTLDAAFEFQNEKGDAPALRVVAESAAAA 345
QY 302 TRLLFNSAPANRVETINPASGP-TESLKDQMGKQWBAEQKAA 343
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 302 TRLLFNSAPANRVETINPASGP-TESLKDQMGKQWBAEQKAA 343
```

```
RESULT 13
US-10-241-742-16
; Sequence 16, Application US/10241742
; Publication No. US20040002147A1
; GENERAL INFORMATION:
; APPLICANT: Short, Jay
; APPLICANT: Weiner, David
; APPLICANT: Chaplin, Jennifer
; APPLICANT: Chi, Ellen
```

```

APPLICANT: Burk, Mark
TITLE OF INVENTION: Nitrilases
FILE REFERENCE: Docket No. US20040014195A1 DIV-013US
CURRENT APPLICATION NUMBER: US/10/440,523
CURRENT FILING DATE: 2003-05-15
PRIORITY APPLICATION NUMBER: US/10/146,772
PRIORITY FILING DATE: 2002-05-15
PRIORITY APPLICATION NUMBER: US 60/309,006
PRIORITY FILING DATE: 2001-07-30
PRIORITY APPLICATION NUMBER: US 60/351,336
PRIORITY FILING DATE: 2002-01-22
PRIORITY APPLICATION NUMBER: US 60/300,189
PRIORITY FILING DATE: 2001-06-21
PRIORITY APPLICATION NUMBER: US 09/751,299
PRIORITY FILING DATE: 2000-12-28
PRIORITY APPLICATION NUMBER: US 60/254,414
PRIORITY FILING DATE: 2000-12-07
PRIORITY APPLICATION NUMBER: US 60/173,609
PRIORITY FILING DATE: 1999-12-29
NUMBER OF SEQ ID NOS: 386
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 16
LENGTH: 348
TYPE: PRT
ORGANISM: Unknown
FEATURE:
OTHER INFORMATION: Obtained from an environmental sample
US-10-440-523-16
Query Match 66.9%; Score 1209; DB 4; Length 348;
Best Local Similarity 67.1%; Pred. No. 4,5e-11;

```

[illegible]

```

; FILE REFERENCE: 09010-900001
; CURRENT APPLICATION NUMBER: US/10/440, 503
; CURRENT FILING DATE: 2003-05-15
; PRIOR APPLICATION NUMBER: US 60/380, 737
; PRIOR FILING DATE: 2002-05-15
; NUMBER OF SEQ ID NOS: 386
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 16
; LENGTH: 348
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Obtained from an environmental sample
; US-10-440-503-16

```

```

Query Match 66.9%; Score 1209; DB 4; Length 348;
Best Local Similarity 67.1%; Pred. No. 4.5e-111;
Matches 230; Conservative 42; Mismatches 69; Indels 2; Gaps 2;

```

```

QY 4 PMTK-YRGAAVQAPVFLDIDRTVEKAIGLIEQAKQDVRLIAPPETWIPGYPPWILGA 62
Db 2 PTKQPFVAAVQAPVFLDIEGAISKGISLIEEASNGAKLIAPEETWIPGYPPWILDS 61
QY 63 PAMGMRVQRYFENSLVRSKQWALADAAARRHGMVVAGYSEBAGSLYMGQALIEGPDG 122
Db 62 PAMGMRVQRYFENSLVRSKQWALADAAARRHGMVVAGYSEBAGSLYMGQALIEGPDG 121
QY 123 DLTAARRKLTPTAERTVFGSDGSHLAVHDTAIGRLGALCWEHIQPLSKYAMYADEQ 182
Db 122 ETIFTRKLTPTAERTVFGSDGSHLAVHDTAIGRLGALCWEHIQPLSKYAMYADEQ 181
QY 183 VHVAVSWPSFLYRGMAVALPEVNTAASQIYAVEGGCYVLASCATVSPENIKVLVDTDPK 242
Db 182 IHIAWSPSFLYRGMAVALPEVNTAASQIYAVEGGCYVLASCATVSPENIKVLVDTDPK 241
QY 243 EMFLKAGGCFAMTFCPGRLAEPLPTEEGGLVADIDLGMIALAKAAADPAGHYSPDV 302
Db 242 EPLLEGGGFTMYGPDGRPLAKPLPENEGGLVADIDLGMIALAKAAADPAGHYSPDV 301
QY 303 TRLLIDRRPAQRVVTLDAAFEFQWEDKGDAPALRVVAESAAA 345
Db 302 TRLLFNGAPANRVEYINPAGSP-TESLKDMGKMGMEAEQOKAA 343

```

Search completed: April 27, 2006, 01:06:08  
 Job time : 113.956 secs

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: April 27, 2006, 01:02:36 ; Search time 18.2372 Seconds  
(without alignments)  
862.816 Million cell updates/sec

Title: US-09-751-299-2

Perfect score: 1806  
Sequence: 1 MSEPTKTRGAAGVAPVFL.....EDKGDPAALRVVAGSAAQAQ 346

Scoring table: BLOSUM62

Gapop 10.0, Gapext 0.5

Searched: 232119 seqs, 45477862 residues

Total number of hits satisfying chosen parameters: 232119

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Published Applications AA New:\*  
1: /SIDS5/prodata/2/pubppaa/US08\_NEW\_PUB\_PEP:\*  
2: /SIDS5/prodata/2/pubppaa/US06\_NEW\_PUB\_PEP:\*  
3: /SIDS5/prodata/2/pubppaa/US07\_NEW\_PUB\_PEP:\*  
4: /SIDS5/prodata/2/pubppaa/US09\_NEW\_PUB\_PEP:\*  
5: /SIDS5/prodata/2/pubppaa/US10\_NEW\_PUB\_PEP:\*  
6: /SIDS5/prodata/2/pubppaa/US11\_NEW\_PUB\_PEP:\*  
7: /SIDS5/prodata/2/pubppaa/US60\_NEW\_PUB\_PEP:\*  
8: /SIDS5/prodata/2/pubppaa/US60\_NEW\_PUB\_PEP:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	869	48.1	356	6	US-10-537-075-7
2	810	44.9	369	6	US-10-919-182-16
3	809	44.8	369	6	US-10-919-182-18
4	807	44.7	369	6	US-10-919-182-4
5	807	44.7	369	6	US-10-919-182-6
6	805	44.6	369	6	US-10-919-182-12
7	804	44.5	369	6	US-10-919-182-14
8	796	44.1	369	6	US-10-919-182-8
9	430.5	22.8	333	7	US-11-096-568A-20687
10	430.5	22.8	351	7	US-11-096-568A-20686
11	399.5	22.1	330	7	US-11-096-568A-6955
12	399.5	22.1	350	7	US-11-096-568A-6954
13	315	17.4	233	7	US-11-096-568A-6956
14	284	15.7	193	7	US-11-096-568A-20688
15	131.5	7.3	901	7	US-11-087-099-2095
16	131.5	7.3	901	7	US-11-188-298-2049
17	110.5	6.1	304	6	US-10-467-657-7410
18	107.5	6.0	532	7	US-11-079-463-5688
19	106.5	5.9	295	7	US-11-079-463-5457
20	105	5.8	259	7	US-11-045-004-822
21	102.5	5.7	336	7	US-11-188-298-20349
22	99.5	5.5	333	7	US-11-188-298-16658
23	95.5	5.3	553	7	US-11-188-298-19842
24	94	5.2	306	7	US-11-188-298-17485
25	93	5.1	3655	7	US-11-075-185-5

26	91.5	5.1	1632	6	US-10-506-454-1147	Sequence 1147, Ap
27	91	5.0	457	7	US-11-059-867-14	Sequence 14, Appl
28	90	5.0	811	7	US-11-200-296B-81	Sequence 81, Appl
29	89.5	5.0	290	7	US-11-188-298-1550	Sequence 1550, Ap
30	89.5	5.0	461	7	US-11-087-039-4404	Sequence 4404, Ap
31	89	4.9	647	6	US-10-915-002-283	Sequence 283, App
32	89	4.9	1192	6	US-10-858-730-72	Sequence 72, Appl
33	88.5	4.9	356	7	US-11-188-298-22086	Sequence 22086, A
34	88.5	4.9	516	7	US-11-096-568A-14918	Sequence 14918, A
35	88.5	4.9	541	7	US-11-096-568A-14917	Sequence 14917, A
36	88.5	4.9	601	7	US-11-096-568A-14916	Sequence 14916, A
37	88	4.9	436	7	US-11-188-298-1186	Sequence 1186, App
38	87.5	4.8	347	7	US-11-188-298-18192	Sequence 18192, A
39	87.5	4.8	492	7	US-11-188-298-1484	Sequence 1484, Ap
40	87	4.8	321	7	US-11-188-298-6685	Sequence 6685, Ap
41	87	4.8	627	7	US-11-188-298-10013	Sequence 10013, A
42	87	4.8	867	7	US-11-079-463-12223	Sequence 12223, Ap
43	86.5	4.8	301	7	US-11-072-512-21830	Sequence 21830, A
44	86.5	4.8	351	7	US-11-096-568A-21830	Sequence 21830, A
45	86.5	4.8	353	7	US-11-096-568A-21829	Sequence 21829, A

## ALIGNMENTS

RESULT 1  
US-10-537-075-7  
; Sequence 7, Application US/10537075  
; Publication No. US20060014291A1  
; GENERAL INFORMATION:  
; APPLICANT: Zelnicki, Maria  
; APPLICANT: Kesseler, Thomas  
; APPLICANT: Hauser, Bernhard  
; TITLE OF INVENTION: L-RHANOSE-INDUCIBLE EXPRESSION SYSTEMS  
; FILE REFERENCE: 12810-00091-US  
; CURRENT APPLICATION NUMBER: US/10/537, 075  
; PRIOR FILING DATE: 2005-06-01  
; PRIOR APPLICATION NUMBER: PCT/EP2003/013367  
; PRIOR FILING DATE: 2003-11-27  
; PRIOR APPLICATION NUMBER: DE 102 56 381.0  
; PRIOR FILING DATE: 2002-12-02  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO: 7  
; LENGTH: 356  
; TYPE: PRT  
; ORGANISM: Alcaligenes faecalis  
US-10-537-075-7

Query Match 48.1%, Score 869, DB 6, Length 356;  
Best Local Similarity 50.3%, Pred. No. 1.5e-71;  
Matches 171; Conservative 54; Mismatches 105; Indels 10; Gaps 2;  
QY 9 RGAAGVAPVLDLDRVEKAIGLEQAKQDVLLAFPTWIPGYFWIMGAPAMQMR 68  
DB 8 RAAVQAASPRYDLAGTGYDKITELARQARDCGLIYGTETWLGYPFHVWLGAPAWLTK 67  
QY 69 FVQRFFENSILVRGSKQWQALDAARRHGMVAVGYSEBAGSLYWGQALFPPDGLIAR 128  
DB 68 YSARYANSLSLDSEAFQRIQAARTLCIFALGYSEBSGSLYIGQLIDKEMLSMR 127  
QY 129 RKLKPTAERVRPREGDSHLAVHDTAIGRGLACWEHIOPLSKYANVYADBOVHVMSV 188  
DB 128 RKLKPTAERVRPREGDSHLAVHDTAIGRGLACWEHIOPLSKYANVYADBOVHVMSV 187  
QY 189 PSFSIYRGMAVALPDEVNTASQIYAVEGCGYVLASCTVSPENIKVLVDPDKEMLPKA 248  
DB 188 PSFSIYRGMAVALPDEVNTASQIYAVEGCGYVLASCTVSPENIKVLVDPDKEMLPKA 247  
QY 249 GGGFAMIFGPDGRALABLPETBEGLLVADIDLMIALAKAADPAGHSRPDVTRLLD 308  
DB 248 GGGFAMIFGPDGRALABLPETBEGLLVADIDLMIALAKAADPAGHSRPDVTRLLD 307

QY 309 ---RRPAQRVTLDAAFEPQNEKGDAPALRVAESAAAA 345  
Db 308 LGHRDPMTRV-----HSKSVTREAEAEQGVQSKIASVA 340

RESULT 2  
US-10-919-182-16

Sequence 16, Application US/10919182  
Publication No. US2006003532A1  
GENERAL INFORMATION:  
APPLICANT: E.I. duPont de Nemours and Company, Inc.  
APPLICANT: Di Cosimo, Robert  
APPLICANT: Payne, Mark  
APPLICANT: O'Keefe, Daniel  
TITLE OF INVENTION: PRODUCTION OF 3-HYDROXYCARBOXYLIC ACID USING NITRILASE MUTANTS  
FILE REFERENCE: C12584 US NA  
CURRENT APPLICATION NUMBER: US/10/919,182  
CURRENT FILING DATE: 2004-08-16  
NUMBER OF SEQ ID NOS: 18  
SOFTWARE: PatentIn version 3.2  
SEQ ID NO 16  
LENGTH: 369  
TYPE: PRT  
ORGANISM: Artificial sequence  
FEATURE:  
OTHER INFORMATION: Mutant nitrilase with Phe168 to Val change  
US-10-919-182-16

Query Match 44.9%; Score 810; DB 6; Length 369;  
Best Local Similarity 46.6%; Pred. No. 3.9e-66;  
Matches 160; Conservative 53; Mismatches 126; Indels 4; Gaps 2;

QY 6 TKYRGAAVQAPVFLDIDRTVEKAIGLIEQAKODVRLAFPEPTWIPGYPFMTWLGAPAW 65  
Db 6 SKFLAATVQAEPPVWLADATIDKISIGIEBAQKASLIAPPEVFIPIGYPMWMLGDVXY 65  
QY 66 GMRPVQRYFENSILVRGSKQWQALADAARRHGMHVAVGSRAGSLYMGQATFGPDGLI 125  
Db 66 SLSTSRKHENSLELGDGRMRRLQLAARRNKIALVMGYSERAGSRYSQVFTIDERGEIV 125  
QY 126 AARRKLKPTHAERTVFEGBDGSRLAVHDTAIGRLGALCWEHIQPLSKYAMVAADEQVHV 185  
Db 126 ANRRKLKPTVERTIYGEINGDTFLTHDFAFGVGGINCEWHQPLSKFMVYSLGEGVHV 185  
QY 186 ASWPSFSLYRGMAVYALGPEVNTAASQIYAVEGCGVYLASCATVSPENIKVLVTPDKEMF 245  
Db 186 ASWPMSPLOPDVQFOLSEANAVTRSYALGQTFVLCSTOVIGPSALFTFCINDEQRL 245  
QY 246 LKAGGPFAMIFGPDGRALAEPLPETEGLLVADIDLGMIATLAKAAADPAGHYSRPDYTRL 305  
Db 246 LPQCGMARIRYGPDSGLAKPLAEDAGIILYABIDLEQIILAKXGADPVGHYSRPDVLVS 305  
QY 306 LLDR---PAQRVTLDAAFEPQNEKGDAPALRVAESAAAA 345  
Db 306 QPDRNHTPVHR-IGIDGRLDVNTSRVENFRLROAEQERQA 347

RESULT 3  
US-10-919-182-18  
Sequence 18, Application US/10919182  
Publication No. US2006003532A1  
GENERAL INFORMATION:  
APPLICANT: E.I. duPont de Nemours and Company, Inc.  
APPLICANT: Di Cosimo, Robert  
APPLICANT: Payne, Mark  
APPLICANT: O'Keefe, Daniel  
TITLE OF INVENTION: PRODUCTION OF 3-HYDROXYCARBOXYLIC ACID USING NITRILASE MUTANTS  
FILE REFERENCE: C12584 US NA  
CURRENT APPLICATION NUMBER: US/10/919,182  
CURRENT FILING DATE: 2004-08-16  
NUMBER OF SEQ ID NOS: 18  
SOFTWARE: PatentIn version 3.2  
SEQ ID NO 18

LENGTH: 369  
TYPE: PRT  
ORGANISM: Artificial sequence  
FEATURE:  
OTHER INFORMATION: Mutant nitrilase with Phe168 to Leu change  
US-10-919-182-18

Query Match 44.8%; Score 809; DB 6; Length 369;  
Best Local Similarity 46.6%; Pred. No. 4.8e-66;  
Matches 160; Conservative 53; Mismatches 126; Indels 4; Gaps 2;

QY 6 TKYRGAAVQAPVFLDIDRTVEKAIGLIEQAKODVRLAFPEPTWIPGYPFMTWLGAPAW 65  
Db 6 SKFLAATVQAEPPVWLADATIDKISIGIEBAQKASLIAPPEVFIPIGYPMWMLGDVXY 65  
QY 66 GMRPVQRYFENSILVRGSKQWQALADAARRHGMHVAVGSRAGSLYMGQATFGPDGLI 125  
Db 66 SLSTSRKHENSLELGDGRMRRLQLAARRNKIALVMGYSERAGSRYSQVFTIDERGEIV 125  
QY 126 AARRKLKPTHAERTVFEGBDGSRLAVHDTAIGRLGALCWEHIQPLSKYAMVAADEQVHV 185  
Db 126 ANRRKLKPTVERTIYGEINGDTFLTHDFAFGVGGINCEWHQPLSKFMVYSLGEGVHV 185  
QY 186 ASWPSFSLYRGMAVYALGPEVNTAASQIYAVEGCGVYLASCATVSPENIKVLVTPDKEMF 245  
Db 186 ASWPMSPLOPDVQFOLSEANAVTRSYALGQTFVLCSTOVIGPSALFTFCINDEQRL 245  
QY 246 LKAGGPFAMIFGPDGRALAEPLPETEGLLVADIDLGMIATLAKAAADPAGHYSRPDYTRL 305  
Db 246 LPQCGMARIRYGPDSGLAKPLAEDAGIILYABIDLEQIILAKXGADPVGHYSRPDVLVS 305  
QY 306 LLDR---PAQRVTLDAAFEPQNEKGDAPALRVAESAAAA 345  
Db 306 QPDRNHTPVHR-IGIDGRLDVNTSRVENFRLROAEQERQA 347

RESULT 4  
US-10-919-182-4  
Sequence 4, Application US/10919182  
Publication No. US2006003532A1  
GENERAL INFORMATION:  
APPLICANT: E.I. duPont de Nemours and Company, Inc.  
APPLICANT: Di Cosimo, Robert  
APPLICANT: Payne, Mark  
APPLICANT: O'Keefe, Daniel  
TITLE OF INVENTION: PRODUCTION OF 3-HYDROXYCARBOXYLIC ACID USING NITRILASE MUTANTS  
FILE REFERENCE: C12584 US NA  
CURRENT APPLICATION NUMBER: US/10/919,182  
CURRENT FILING DATE: 2004-08-16  
NUMBER OF SEQ ID NOS: 18  
SOFTWARE: PatentIn version 3.2  
SEQ ID NO 4  
LENGTH: 369  
TYPE: PRT  
ORGANISM: Acidovorax facilis 72W  
US-10-919-182-4

Query Match 44.7%; Score 807; DB 6; Length 369;  
Best Local Similarity 46.6%; Pred. No. 7.3e-66;  
Matches 160; Conservative 52; Mismatches 127; Indels 4; Gaps 2;







US-11-096-568A-20686

Query Match 23.8%; Score 430.5; DB 7; Length 351;

Best Local Similarity 35.0%; Pred. No. 1.6e-31;

Matches 117; Conservative 52; Mismatches 128; Indels 37; Gaps 8;

```
QY 6 TTKRGAAVOAARVFLDLDRTEKALIGLIEQAKQDVRLLIAPETWIPGYP---FWIWIG 61
DB 28 TTAATVTVQASSVFDYTPATLDKAEKLVAAAGYSQVLVFEVFGVYFGHSTFGLVVG 87
QY 62 -APANGKRFVQRYFENSLVRGSKQWQALDAAARRHGMHVAGYSERAGSLYMGQALFSP 120
DB 88 NRTAGKEDFQKXHSALIDVPGPEVSRSLBALAGKYFVLIVGVERAGTYLVNTVLSFDP 147
QY 121 DGDLLAARKKLTPTAERTVFGEGDSHLAVHDTAIGRLGALCCWEHIOPISKVMTAAD 180
DB 148 LKRYIGKHKWPTALERVFWFGDSTIPVYDTPIGKKGALICWENRMLRTAMAYAG 207
QY 181 EGVHVASWPSFLYRGMAYALGPEVNTAASQI-----VAVEGQCYVLASCA-----TV 228
DB 208 IEIYCA-----PTVDCMPTWLSMTHIALBEGCFVLSACQFCRKNRYP 251
QY 229 SPEWIKV-LVDTPEKEMFLKAGGPFAMIFGPDGRALAEPLPETEGLLVADIDLGMIALA 287
DB 252 PPEYTFCGLEEPSDESVCSSG--SVIISPLGTVLAGPNYS-EXLLTADLDLGEIVRA 308
QY 288 KAAADPAGHYRSDVTRILLDRRPAQRVYTLDA 321
DB 309 KFDVGVGHYSRPEVLSLVVKSDFKRAVSIFSA 342
```

RESULT 11

US-11-096-568A-6955

; Sequence 6955; Application US/11096568A

; Publication No. US20060048240A1

; GENERAL INFORMATION:

; APPLICANT: Alexandrov, Nikolai et al.

; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides

; FILE REFERENCE: 2750-1592PUS2

; CURRENT APPLICATION NUMBER: US/11/096,568A

; CURRENT FILING DATE: 2005-04-01

; NUMBER OF SEQ ID NOS: 34471

; SEQ ID NO 6955

; LENGTH: 330

; TYPE: PRT

; ORGANISM: Glycine max

; FEATURE:

; NAME/KEY: misc feature

; LOCATION: (1)..(330)

; OTHER INFORMATION: Ceres Seq. ID no. 15169318

US-11-096-568A-6955

Query Match 22.1%; Score 399.5; DB 7; Length 330;

Best Local Similarity 33.3%; Pred. No. 1e-28;

Matches 112; Conservative 57; Mismatches 136; Indels 31; Gaps 11;

```
QY 9 GGAATVQAAARVFLDLDRTEKALIGLIEQAKQDVRLLIAPETWIPGYP---FWIWIG-AP 63
DB 11 RATVVOASTIIFYDTPATLDKAEKLVAAAGYSQVLVFEVFGVYFGHSTFGLVVG 70
QY 64 AMGKRFVQRYFENSLVRGSKQWQALDAAARRHGMHVAGYSERAGSLYMGQALFSPDG 123
DB 71 VKGREPFKXHSALIDVPGPEVDRLLAAMAGKXKVLWGVIERDGYTLVCTVLPFDSQGH 130
QY 124 LIAARRKLTPTAERTVFGEGDSHLAVHDTAIGRLGALCCWEHIOPISKVMTAAD 183
DB 131 YLGRKRIKIMPTALERVFWFGDSTIPVYDTPIGKKGALICWENRMLRTAMAYAGVEI 190
QY 184 HVASWPSFLYRGMAYALGPEVNTAASQIYAVEGQCYVLAS---CA---TVSPEMI--K 234
DB 191 YCA--PT-----ADARDVWQASMTHTIALBEGCFVLSANQFCRRRDYPPPEYFVAG 239
```

```
QY 235 VLVD--TPDKEMFLKAGGPFAMIFGPDGRALAEPLPETEGLLVADIDLGMIALAKAADP 293
DB 240 TEVDLTPDS--VVCAGG--SVIISPLGAVLAGPNVD-GEALISADLDLGEIARAKFDFV 294
```

```
QY 294 AGHYRSDVTRILLDRRPAQRVYTLDAAPFQNEDEK 329
DB 295 VGHYSRPEVLSLVVKSDFKRAVSIFSA 330
```

RESULT 12

US-11-096-568A-6954

; Sequence 6954; Application US/11096568A

; Publication No. US20060048240A1

; GENERAL INFORMATION:

; APPLICANT: Alexandrov, Nikolai et al.

; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides

; FILE REFERENCE: 2750-1592PUS2

; CURRENT APPLICATION NUMBER: US/11/096,568A

; CURRENT FILING DATE: 2005-04-01

; NUMBER OF SEQ ID NOS: 34471

; SEQ ID NO 6954

; LENGTH: 350

; TYPE: PRT

; ORGANISM: Glycine max

; FEATURE:

; NAME/KEY: misc feature

; LOCATION: (1)..(350)

; OTHER INFORMATION: Ceres Seq. ID no. 15169317

US-11-096-568A-6954

Query Match 22.1%; Score 399.5; DB 7; Length 350;

Best Local Similarity 33.3%; Pred. No. 1.1e-28;

Matches 112; Conservative 57; Mismatches 136; Indels 31; Gaps 11;

```
QY 9 GGAATVQAAARVFLDLDRTEKALIGLIEQAKQDVRLLIAPETWIPGYP---FWIWIG-AP 63
DB 31 RATVVOASTIIFYDTPATLDKAEKLVAAAGYSQVLVFEVFGVYFGHSTFGLVVG 90
QY 64 AMGKRFVQRYFENSLVRGSKQWQALDAAARRHGMHVAGYSERAGSLYMGQALFSPDG 123
DB 91 VKGREPFKXHSALIDVPGPEVDRLLAAMAGKXKVLWGVIERDGYTLVCTVLPFDSQGH 150
QY 124 LIAARRKLTPTAERTVFGEGDSHLAVHDTAIGRLGALCCWEHIOPISKVMTAAD 183
DB 151 YLGRKRIKIMPTALERVFWFGDSTIPVYDTPIGKKGALICWENRMLRTAMAYAGVEI 210
QY 184 HVASWPSFLYRGMAYALGPEVNTAASQIYAVEGQCYVLAS---CA---TVSPEMI--K 234
DB 211 YCA--PT-----ADARDVWQASMTHTIALBEGCFVLSANQFCRRRDYPPPEYFVAG 259
QY 235 VLVD--TPDKEMFLKAGGPFAMIFGPDGRALAEPLPETEGLLVADIDLGMIALAKAADP 293
DB 260 TEVDLTPDS--VVCAGG--SVIISPLGAVLAGPNVD-GEALISADLDLGEIARAKFDFV 314
QY 294 AGHYRSDVTRILLDRRPAQRVYTLDAAPFQNEDEK 329
DB 315 VGHYSRPEVLSLVVKSDFKRAVSIFSA 350
```

RESULT 13

US-11-096-568A-6956

; Sequence 6956; Application US/11096568A

; Publication No. US20060048240A1

; GENERAL INFORMATION:

; APPLICANT: Alexandrov, Nikolai et al.

; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides

; FILE REFERENCE: 2750-1592PUS2

; CURRENT APPLICATION NUMBER: US/11/096,568A

; CURRENT FILING DATE: 2005-04-01

; NUMBER OF SEQ ID NOS: 34471

; SEQ ID NO 6956





Db 49 ValPheIleProGlyTyrProTyrHisIleThrValAspSerProLeuAlaGlyMetAla 68  
Qy 193 CAATTGTAGCCCAATACCATAGAACTCATTTGATGGATGGCCCTCAAGTAAAGCGC 252  
Db 69 LysPheAlaValArgTyrHisIleGlnAsnSerIleuThrMetAspSerProHisValGlnArg 88  
Qy 253 ATTTCAGATGACGCAAGCGGTTGGAACTATGTCACCCCTGGGGATAGTGAACGGGTC 312  
Db 89 LeuLeuAspAlaAlaArgAspHisAsnIleAlaValAlaGlyIleSerGlnArgAsp 108  
Qy 313 GGTGACACCCCTTTTCATGACGTAGGTTGATAGCGGTAATATGGACACCATATGGGGCC 372  
Db 109 GlyIleSerLeuTyrMetThrGlnLeuValIleAspAlaAspGlyGlnLeuAlaArg 128  
Qy 373 CGGCGAAAGTTGAACCTACTTTTGTGTGAACGTATTTGTTGGCGAAGGGATGGTTCA 432  
Db 129 ArgArgLysLeuLysProThrHisValGlnArgSerValTyrGlyGlnGlyAsnGlySer 148  
Qy 433 TCGCTAGCGGTTTTCGAGACGTCGTGGAAAGCGTGGTGCTTATGCTGTGGAGAC 492  
Db 149 AspIleSerValTyrAspMetProPheAlaArgLeuGlyAlaLeuAsnGlyTrpGlnHis 168  
Qy 493 CTTGACCGCTTAACAAATAAGCTTTGTATGACAAATAAGAAATTCATTTGCGGCT 552  
Db 169 PheGlnThrLeuThrLysTyrAlaMetTyrSerMetHisGlnGlnValHisValAlaAsp 188  
Qy 553 TGGCGGAGCTTATGACCTTATCTTAATGCGGCAAGCGCTGGGGCTGATGTCAATGA 612  
Db 189 TrpProGlyMetSerLeuTyrGlnProGlyValProAlaPheGlyValAspAlaGlnLeu 208  
Qy 613 GCGGCTCTGCAATCTATAGCGGTTGAAGGCAATGCTTACACTAGCGCTGGAGCGCTC 672  
Db 209 ThrAlaThrArgMetTyrAlaLeuGlnGlyGlnThrPheValValCysThrThrGlnVal 228  
Qy 673 GTTTCACAAATCATGATGATGCTTTGTACAGATGAAGAAAGCATGCGTTGGCTGTG 732  
Db 229 ValThrProGlyAlaHisIleGlnPhePheCysAspAsnAspArgGlnArgLysLeuIleLys 248  
Qy 733 GCTGGTGTGACACTACGATATCATAGGCGCTGATGGTGTGACTGTGTGGCGCTTT 792  
Db 249 ArgGlyGlyLysPheAlaArgIleIleGlyProAspArgLysAspLeuAlaThrProLeu 268  
Qy 793 GCCGAAATGAAGGGATTTCTGACGCAAACTGATCTCGAGATGCGATCTTGCT 852  
Db 269 AlaGluAspGlyGlnGlyIleLeuTyrAlaAspIleAspLeuSerAlaIleThrLeuAla 288  
Qy 853 AAAATGGCGGAGACCTGTGCTATTTATTCCTCCGACATTAATCGCTTGCTAATA 912  
Db 289 LysGlnAlaAlaAspProValGlyHisTyrSerArgProAspValLeuSerLeuAsnPhe 308  
Qy 913 GATCGGAGCCTTAATTCGGTA 936  
Db 309 AsnGlnArgHisThrThrProVal 316

RESULT 2  
A47181  
nitriIase (EC 3.5.5.1), arylacetone-specific - Alcaligenes faecalis  
C/Species: Alcaligenes faecalis  
C/Date: 21-Sep-1993 #sequence\_revision 18-Nov-1994 #text\_change 31-Dec-2004  
C/Accession: A47181; S13860  
R/Kobayashi, M.; Izui, H.; Nagasawa, T.; Yamada, H.  
Proc. Natl. Acad. Sci. U.S.A. 90, 247-251, 1993  
A/Title: NitriIase in biosynthesis of the plant hormone indole-3-acetic acid from indole  
A/Reference number: A47181; MUID:93126352; PMID:8419930  
A/Contents: JMJ3  
A/Accession: A47181  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-356 <KOB>  
A/Cross-references: UNIPROT:P20960; UNIPARC:UP10000033514; GB:D13419; NID:G216202; PIDN:  
A/Note: sequence extracted from NCBI Backbone (NCBIN:122081, NCBI:P122082)  
R/Nagasawa, T.; Mauger, J.; Yamada, H.

Eur. J. Biochem. 194, 765-772, 1990  
A/Title: A novel nitriIase, arylacetone-nitriIase, of Alcaligenes faecalis JMJ3. Purification  
A/Reference number: S13860; MUID:91099356; PMID:2269298  
A/Accession: S13860  
A/Status: preliminary  
A/Molecule type: protein  
A/Residues: 1-33 <NAG>  
A/Cross-references: UNIPARC:UP10000175E31  
A/Superfamily: nitriIase (carbon-nitrogen hydrolase)  
C/Keywords: hydrolase

Alignment Scores:  
Pred. No.: 2,77e-69 Length: 356  
Score: 819.00 Matches: 156  
Percent Similarity: 66.9% Conservative: 54  
Best Local Similarity: 49.7% Mismatches: 104  
Query Match: 44.9% Indels: 0  
DB: Gaps: 0

US-09-751-299-3 (1-1014) x A47181 (1-356)

Qy 4 AAAGAGCTTATCAAGTGCCTGCGTCAAGCCGCCCATCTTACATGATTTGGAGCG 63  
Db 4 ArgLysIleValAlaArgAlaAlaValAlaGlnAlaIleAspProAsnTyrAspLeuAlaThr 23  
Qy 64 ACGGTGACAAACCATTTGATGATGGAAGAGCAGACAGTAAATATGCTCGTGTATC 123  
Db 24 GlyValAspLysThrIleGlyLeuAlaArgGlnAlaArgAspGlyGlyCysAspLeuIle 43  
Qy 124 GCGTTTCGGAACCTTGATTCAGAGTCCAGCTACCATGCTTTCTTGGCTGATCCAGCA 183  
Db 44 ValPheGlyLysThrTrpLeuProGlyTyrProPheHisValTrpLeuGlyAlaProAla 63  
Qy 184 TGGCAATGCAATTGTGAACCCCAATACCATGAAGAATCTGAGTTGATGGCCCTCAA 243  
Db 64 TrpSerLeuLysTyrSerAlaArgTyrTyrAlaAsnSerLeuSerLeuAspSerAlaGlu 83  
Qy 244 GCTAATGCAATTCAGATGCGAGCCAGCGGTTGGAGATTCATGTCACCTGGGGATAGT 303  
Db 84 PheGlnArgIleAlaGlnAlaAlaArgThrLeuGlyIlePheIleAlaLeuGlyTyrSer 103  
Qy 304 GAACGGGTGGTGGACACCTTTACATGATGATGATGCTTATGAGCGATATATGGTGAACC 363  
Db 104 GlnArgSerGlyGlySerLeuTyrLeuGlyGlnCysLeuIleAspArgLysGlyGlnMet 123  
Qy 364 ATTGGGCGCCGCGCAAGCTTGAACCTACTTTGTGAAGTACTTGTTCGGGGAAGG 423  
Db 124 LeuTrpSerArgArgLysLeuLysProThrHisValGlnArgThrValPheGlyGlnGly 143  
Qy 424 GATGATTCATGCGTAGCGGTTTTTCGAGACGTCTGTTGAAAGGCTGGGCTTATGCTGT 483  
Db 144 TyrAlaArgAspLeuIleValSerAspThrGlnLeuGlyArgValGlyAlaLeuCysGly 163  
Qy 484 TGGAGACACTTCAACCGCTTACAAATATGCTTTGTATGCAAAATGAAGATTTAT 543  
Db 164 TrpGlnHisLeuSerProLeuSerIleTyrAlaLeuTyrSerGlnHisGlnAlaIleHis 183  
Qy 544 TGTGCGGCTTGGCGGAGCTTATGAGCTTATCTTAATGCGGGAAGCCCTGGGCGTGTAT 603  
Db 184 IleAlaAlaTrpProSerPheSerLeuLysSerGlnAlaHisValAlaLeuSerAlaLys 203  
Qy 604 GTCAATGAGCGGCTCTGCAATCTATGCGTTGAAGGCAATGCTTGTATGCGCTG 663  
Db 204 ValAsnMetAlaAlaSerGlnIleTyrSerValGlnGlyGlnCysPheThrIleAlaAla 223  
Qy 664 TGTGCGCTGTTTCACAATCATGATGATGATGCTTTGTACAGATGACGAAAGCATGCG 723  
Db 224 SerSerValValThrGlnGlnThrLeuAspMetLeuGlnValGlyGlnHisAsnAlaSer 243  
Qy 724 TTTGCTTTCGCTGTGGTGGGACACTCACTGATTCATAGGCGCTGATGGTGTGATGCTG 783  
Db 244 LeuLeuLysValGlyGlyGlySerSerMetIlePheAlaProAspGlyArgThrLeuAla 263



Db 27 LysThrCysIysLeuIleAlaGluAlaIleSerMetGlyAlaIysValIleGlyPhePro 46  
QY 133 GAAAGTGGATTCAGAGCTACCCAGTGTTCCTTGGCTTGA- 177  
Db 47 GluAlaPheIleProGlyTyrProGlyTyrIleThrIleThrSerAspMetAspPheThrGly 66  
QY 178 CCAGATGGGCGAATGTCATTTGTACGCCAATACATGAACTGATGAGTGGATGGC 237  
Db 67 MetMetIlePheAlaValLeuPhe-----LysAsnAlaIleGluIleProSer 81  
QY 238 CCTCAAGCTAAGCGCATTTCAATGACATGACCAAGCGGTTGGAAATCATGATGACCCCTGGG 297  
Db 82 LysGluValGlnGlnIleSerAspAlaIaIysLysAsnGlyAlaItyrValCysValSer 101  
QY 298 ATGAGTGAACGGGTCCGGTGGACCCCTTACATCACTGCTGTGATGAGGATTAATGCT 357  
Db 102 ValSerGluIysAspAsnAlaSerLeuTyrIleThrGlnLeuThrPheAspProAsnGly 121  
QY 358 GACACCATTTGGGGCCCGGAAAGTTGAACCTACTTTTGTGAACGTACTTTTGGC 417  
Db 122 AsnLeuIleGlyLysIleAlaGlySerPheIleProThrSerSerGluAlaValIleProGly 141  
QY 418 GAAGGCGATGCTTCATCCGCTTACGAGCTTTTCAAGACGTCTGTGAAAGCTGGTGGCTTA 477  
Db 142 AspGlyAspGlySerMetAlaProValPheIysThrGluTyrGlyAsnLeuGlyGlyLeu 161  
QY 478 TGCCTTTGGGAGACCTTCAACCGCTTACAAATATGCGTTGTATGCAAAATGAAG 537  
Db 162 GlnCysThrGluIleAsnLeuProIleAsnIleAlaIleMetGlySerLeuAsnGln 181  
QY 538 ATTCATTTGGCGGCTTGGCCGAGCTTTAGCCTTATTCCTAATGCGGCAAGCCCTGGG 597  
Db 182 ValHisValAlaSerThrProAlaPhe-----ValProLysGlyAlaValSerSerArg 199  
QY 598 CCTGATGTCATGTACGGCC-----TCTCGAATCAT 630  
Db 200 ValSerSerValCysAlaSerThrAsnAlaMetHisGlnIleIleSerGlnPheTyr 219  
QY 631 GCCGTGAAGGCAATGCTGCTGACTAGCGTCTGTGCGCTGCTTCAAAATCCATGATC 690  
Db 220 AlaIleSerAsnGlnValIleValIleMetSerThrAsnLeuValGlyGlnAspMetIle 239  
QY 691 GATATGCTTTGTACGATGACGAAGAAAGCATGCGTCTTCTGCGCTGGTGGTGGACACTCA 750  
Db 240 AspMetIleGlyLysAspGluPheSerLysAsnPheLeuProLeuGlySerGlyAsnThr 259  
QY 751 CGTATCATAGGCGCTGATGCTGATGCTGCGCGCGCTTGGCCGAATAAGAGAGGAT 810  
Db 260 AlaIleIleSer--AsnThrGlyIleIleLeuAlaSerIleProGlnAspAlaGluGly 278  
QY 811 ATTCTTACGCAACCTTGATCTCTGAGTACGATCCTTGTCTTAAATGCGCGACACCT 870  
Db 279 IleAlaValAlaGluIleAspLeuAsnGlnIleIleTyrGlyLysIleThrLeuAspPro 298  
QY 871 GCTGCTCATTTTCCCGTCCGACATTAATGCTGCTTGTCTTAAATGAGCCCTTAATTA 930  
Db 299 AlaGlyHisTyrSerThrProGlyPheLeuSerLeuThrPheAspGlnSerGlnIleVal 318  
QY 931 CCGAGTACTGAAT 945  
Db 319 ProValLysLysIle 323  
RESULT 5  
A28658  
nitriIase (BC 3.5.5.1) - Klebsiella ozaenae  
C/Species: Klebsiella ozaenae  
C/Date: 10-Sep-1999 #revision 10-Sep-1999 #text\_change 31-Dec-2004  
C/Accession: A28658  
R/Strainer, D.M.; Malyj, L.D.; McBride, K.E.  
J. Biol. Chem. 263, 6310-6314, 1988  
A/Title: Purification and properties of a nitriIase specific for the herbicide bromoxynil  
A/Reference number: A28658; MUID:88198177; PMID:2834373  
S A/Accession: A28658

A/Molecule type: DNA  
A/Residues: 1-349 <Str>  
A/Cross-references: UNIPROT:P10045; UNIPARC:UPI000013051A; GB:J03196; NID:G149174; PIDN:J  
C/Superfamily: nitriIase (carbon-nitrogen hydrolase)  
C/Keywords: hydrolase  
Alignment Scores:  
Pred. No.: 8 71e-52 Length: 349  
Score: 634.50 Matches: 144  
Percent Similarity: 59.0% Conservative: 52  
Best Local Similarity: 43.4% Mismatches: 129  
Query Match: 34.7% Indels: 7  
Gaps: 3  
US-09-751-299-3 (1-1014) x A28658 (1-349)  
QY 1 ATGAAGAAGACTATCAAGTGCCTGGTGCAGCCGCCGATCTACATGATTTGGAG 60  
Db 1 MetAspThrThrPheIysAlaAlaAlaValGlnAlaGluProValIlePheMetAspAlaAla 20  
QY 61 GCGAGCGTGCACAAACCATTTGATGAGTGAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 120  
Db 21 AlaThrAlaAspLysIleValThrLeuValAlaLysAlaAlaAlaAlaGlyAlaGlnLeu 40  
QY 121 ATCCCTTTCCGGAACCTTGATTCAGGCTTACAGCTACCATGCTTCTTGGCTTGAACCA 180  
Db 41 ValAlaPheProGluLeuThrIleProGlyTyrProGlyPheMetLeuThrIleAsnGln 60  
QY 181 GCATGGGCATGCAATTTGTACGCCCATATCCATAGAAATCATGAGACTTGGAGTGGCCCT 240  
Db 61 ThrGluThrLeuProPheIleIleLysTyrArgLysGlnAlaIleAlaAlaAspGlyPro 80  
QY 241 CAAGCTAAGCGCATTTGATGAGTGCAGCAAGCGGTTGGAAATCATATGTCACCTGGAGATG 300  
Db 81 GluIleGluIysIleAlaGlyCysAlaAlaGlnGlnIleAsnIleAlaLeuSerPheGlyTyr 100  
QY 301 AGTGAAGCGGTGGTGGACCCCTTACATCATGATGCTTCAATGCGCATTAATGCTGAC 360  
Db 101 SerGluIleAlaGlyAlaGlyThrLeuTyrMetSerGlnMetLeuIleAspAlaAspGlyLe 120  
QY 361 ACCATTTGGGCGCGCGCAAGTTGAAACCTTCTTGTGAAACGTACTTGTGGCGCA 420  
Db 121 ThrIysIleAlaGlyAlaGlyLysLeuIleProThrArgPheGluIleGluLeuPheGlyGlu 140  
QY 421 GGGAGTGGTATGCTAGCGGTTTTCGAGACGTCTGTGGAAGCGTGGTCTTATGC 480  
Db 141 GlyAspGlySerAspLeuGlnValAlaGlnThrSerValGlyArgValGlyAlaLeuAsn 160  
QY 481 TGTGGGAGCACCTTCAACCGCTTAACAAATACCTTTGTATGACAAATGAAGAGATT 540  
Db 161 CysAlaGluAsnLeuGlnSerLeuAsnLysPheAlaLeuAlaIleGluGlyGlnGlnIle 180  
QY 541 CATTTGGCGGCTTGGCGAGCTTATGACTTATCTTAATGCGCGAAGCCCTGGGCGCT 600  
Db 181 HisIleSerAlaIlePro--PheThrLeu-----GlySerProValLeuValGly 196  
QY 601 GATGTCATGTAGCGGCGCTCTGAATCTATGCGCTTGAAGGCAATGCTTGTACTAGCG 660  
Db 197 AspSerIleGlyAlaIleAsnGlnValTyrAlaIleGluIleThrGlyThrPheValLeuMet 216  
QY 661 TCGTGTGGCGCTGTTTCAATGATGATGATGCTTGTATGAGTGAAGAAAGCAT 720  
Db 217 SerThrGlnValValGlyProThrGlyIleAlaAlaPheGluIleGluAspArgTyrAsn 236  
QY 721 GCGTTCCTTGGCTGTGTGTGACACTACAGTATCATAGGCGCTGATGTGTGACTTG 780  
Db 237 ProAsnGlnTyrLeuGlyGlyGlyTyrAlaArgIleTyrGlyProAspMetGlnLeuLys 256  
QY 781 GTCCGCGCTCTTGGCGAAGAAAGAGGATTTCTTACGAAACCTTGATCTGAGATA 840  
Db 257 SerLysSerLeuSerProThrGluGlnGlyLeuValTyrAlaGluIleAspLeuSerMet 276  
QY 841 CGCATCTTGTCTTAAATGCGCGAGACCCCTGCTGCTATTAATCCGTCGCGACATTAAT 900



Db 277 LeuGLuLaLaLysTYrSerLeuAspProThrGlyHisTYrSerArgProAspValPhe 296

QY 901 CGCTTGCTAAATAGATCGCAGCCCTTAATTACCGGAGTGTGAAT-----GAAGGT 951

Db 297 SerValSerIleAsnArgGlnArgGlnProAlaValSerGluValIleAspSerAsnGly 316

QY 952 GATCTTCGTCTTACGCTTTGGGTAAAGCGCTGTAG 987

Db 317 AspGluAspProArgAlaAlaCysGluProAspGlu 328

RESULT 6

QJ01613

cyanide hydratase (EC 4.2.1.66) - imperfect fungus (Gloeocercospora sorgh1)

C/Species: Gloeocercospora sorgh1

C/Date: 30-Sep-1993 #sequence\_revision 30-Sep-1993 #text\_change 31-Dec-2004

C/Accession: QJ01613

R:Wang, P.; VanEtten, H.D.

Biophys. Res. Commun. 187, 1048-1054, 1992

A>Title: Cloning and properties of a cyanide hydratase gene from the phytopathogenic fun

A/Reference number: QJ01613; MUID:92412068; PMID:1382413

A/Accession: QJ01613

A:Molecule type: DNA

A:Residues: 1-368 <MAN>

A/Cross-references: UNIPROT:P32964; UNIPARC:UPI0000128C2E; GB:M99044; NID:G168196; PIDN

C/Comment: This enzyme converts HCN to formamide.

C/Genetics:

A:Gene: Chc

C:Superfamily: nitrilase (carbon-nitrogen hydrolase)

C/Keywords: carbon-oxygen lyase; glycoprotein; hydro-lyase

F:286/Binding site: carbonyl hydrate (asn) (covalent) #status predicted

Alignment Scores:

Pred. No.: 1.28e-44 Length: 368

Score: 559.00 Matches: 128

Percent Similarity: 56.1% Conservative: 56

Best Local Similarity: 39.08 Mismatches: 118

Query Match: 30.6% Indels: 26

DB: 2 Gaps: 7

US-09-751-299-3 (1-1014) x QJ01613 (1-368)

QY 16 AAGTCGCTGCGTGGCAAGCCGCCGATCTACATGATTTGGAGCGCAGTGAACAA 75

Db 7 LysAlaLaValValThrSerGluProValTyrGluAsnLeuGluGlyValValLys 26

QY 76 ACCATGATGTGATGAGGAAGCAGCAGCATATATGTCGTCTGATGCGCTTCGCCGAA 135

Db 27 ThrIleGluPheIleAsnGlnAlaGlyLysAlaGlyCysValLeuIleAlaPheProGlu 46

QY 136 ACTTGATTCAGAGCTACCCATGGTTGCTTCTTGCTTGATCTGACACGACGATGGCAATGAA 195

Db 47 ValTrpIleProGlyTyrProTyrTrpMetTrpLysValAsnTyrLeuGlnSerLeuPro 66

QY 196 TTGTGACGCCAATAACATGAGAACTCATTTGAGTTGATGGCCCTCAAGCTTAAGCATT 255

Db 67 MetLeuLysAlaTyrArgGluAsnSerIleAlaMetAspSerSerGluMetArgArgIle 86

QY 256 TCAGATGAGCGCAAGCGGTGGGAATCATGTTCACCTCGGGGAGAGTGAAGCGGTCGT 315

Db 87 ArgAlaAlaAlaAspAsnGlnIleTyrAlaSerIleGlyValSerGlnIleAspHis 106

QY 316 GGCAACCTTTACATCAGTCAAGTGGTTCATAGGACATATGATGTCACACATTTGGGCCCG 375

Db 107 AlaThrLeuTyrLeuThrGlnValLeuIleSerProLeuGlnAspValIleAsnHisArg 126

QY 376 CGAAAGTTGAACCTACTTTTGTGAACGTACTTTGTCGGCGAAGGGAGTGTTCATCG 435

Db 127 ArgLysIleLysProThrHisValGlnLysLeuValTyrGluAspGlySerGluAspSer 146

QY 436 CTA---GGCGTTTCGAGCGCTGTTGGAAAGCGTGGCGCTTATGCTGTGGGAACAC 492

Db 147 PheGluProValThrGlnThrGlnIleLysArgLeuGlnLysLeuAsnCysTrpGluAsn 166

QY CTTCAACCGCTTAAACAATAATGCGTTGTATGCAGCAAAATGAAGAATTGTCGGCGT 552  
::: ||||| . ||||| : ||||| :: ||||| :: ||||| |||||  
Db MetanProPheueuYSerieulaValaAlaArgIgluGIInIIehisValaIala 186  
553 TGGCCGAGCCTTAAGCCTTATCTTAATGCGCGCAAGCCCGGGGCGCTGAT-----GTC 606  
||||| : ||||| :: ||||| :: ||||| :: |||||  
Db TrpPro-----ValTrProAspLeuSerIylselnValHisProAspProAlaThr 203  
607 AATGTAAGCG-----GCCCTTGAAATTATGCCGTTGAAGGCAATGC 648  
||| ||| : : :  
Db AsenTryalLaaspProAlaseAseleuValThrProAlaTryalIleGluthrGlyThr 223  
649 TTGCGACTAGCGCGTGGTGGCGTGGCTGTTTCACATCCATGATGCANATGCTTTGT----- 702  
:::||||| ||||| : ||||| : ||||| : |||||  
Db TrpValLeuAlaProPheGlnArgIleSeValGluGlyLeuLysValArgHisThrProPro 243  
224 TrpValLeuAlaProPheGlnArgIleSeValGluGlyLeuLysValArgHisThrProPro 243  
703 -----ACAGTGAAGCAAAAGCATGCTGTGCTTCGGCGTGGTGAAGAC 747  
||| ||| : ||| : :  
Db GlyValGluProGluThrAspAlaThrProTyR-----AsnGlyHis 257  
244 GlyValGluProGluThrAspAlaThrProTyR-----AsnGlyHis 257  
QY TCACGTATCATGAGGCGCTGATGCTGATGCTGATGCTGCGCGCTTGGCGAAAATGAAG 807  
:::||||| : ||||| ||||| ||||| : : :  
Db AlaArgIlePheArgProAsp--GlySerLeuTryAlaLysProAlaValaAspPheAsp 276  
258 AlaArgIlePheArgProAsp--GlySerLeuTryAlaLysProAlaValaAspPheAsp 276  
QY GGTATTTCTGTAGCAAAACCTGATCCTGAGAGTAGCATCCTGTGTAATAATGCGCGCAGAC 867  
||||| : ||||| : ||||| ||||| ||||| |||||  
Db GlyLeuMetTryValaAspIleAspLeuAsnGluSerHisleuThrLysAlaLeuAlaAsp 296  
277 GlyLeuMetTryValaAspIleAspLeuAsnGluSerHisleuThrLysAlaLeuAlaAsp 296  
868 CCTGTGTGTCATTATTCCTCCGTCGCGACATTACTGCTTGTGTAATAGATCGGACCCCTAAA 927  
||||| ||||| ||||| ||||| : ||||| : |||||  
Db PheAlaGlyHisIstyrtetnArgProAspLeuIlleArgLeuLeuValaSprThrArgArgLys 316  
297 PheAlaGlyHisIstyrtetnArgProAspLeuIlleArgLeuLeuValaSprThrArgArgLys 316  
QY 928 TTACCGGTGTGGAATTTGAAGGT 951  
||| ||| : ||| : |||  
Db GluLeuValThrGluValGlyGly 324  
317 GluLeuValThrGluValGlyGly 324

RESULT 7  
S77025  
nitriIase (EC 3.5.5.1) - Synechocystis sp. (strain PCC 6803)  
NAlternate names: protein e110784  
CISpecies: Synechocystis sp.  
A.Variety: PCC 6803  
C.Date: 25-Apr-1997 #sequence\_revision 25-Apr-1997 #text\_change 31-Dec-2004  
R.Accession: S77025  
R.Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asemizu, E.; Nakamura, Y.; Miyajima, N.;  
O. K.; Okumura, S.; Shimpō, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasuda  
DNA Res. 3, 109-136, 1996  
A>Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocystis  
s.  
A.Reference number: S74322; PMID:97061201; PMID:8905231  
A.Accession: S77025  
A>Status: nucleic acid sequence not shown; translation not shown  
A.Molecule type: DNA  
A.Residues: 1-346 <KAN>  
A.Cross-references: UNIPROT:O55949; UNIPARC:UPI00000D7133; EMBL:D64005; GB:AB001339; NID  
A.Note: The nucleotide sequence was submitted to the EMBL Data Library, June 1996  
C.Name(s):  
C.Genetic(s):  
A.Start codon: GTG  
C.Superfamily: nitriIase (carbon-nitrogen hydrolase)  
C.Keywords: hydrolase

Alignment Scores:

Pred. No.:	Length:
2_93e-41	346
Score: 523.50	Matches: 109
Percent Similarity: 58.0%	Conservative: 68
Best Local Similarity: 35.7%	Mismatches: 119
Query Match: 28.7%	Indels: 9
DB: 2	Gaps: 4

US-09-751-289-3 (1-1014) x S77025 (1-346)

13 ATCAAAGTCGCTGCGTCGACAGCCGCCGCGATTTCAATGATTTGGAGGCGACGCTGAC 72

[illegible]

A:Accession: T49147  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-346 <DNA>  
A:Cross-References: UNIPROT:P32961; UNIPARC:UPI0000001AE2; EMBL:AL353865; GSPDB:GN00061.;  
A:Experimental source: cultivar Columbia; BAC clone T10D17  
R.Hillebrand, H.; Bartling, D.; Weiler, E.W.  
Plant Mol. Biol. 36, 89-99, 1998  
A>Title: Structural analysis of the n1c2/n1c1/n1c3 gene cluster encoding nitrilases, enz  
A:Reference number: Z24515; NCID:8022831; PMID:9484465  
A:Accession: T52260  
A:Status: preliminary; translated from GB/EMBL/DBD  
A:Molecule type: DNA  
A:Residues: 1-346 <HTL>  
A:Cross-References: UNIPARC:UPI0000001AE2; EMBL:Y07648; PIDN:CAAE8935.2  
A:Experimental source: cultivar Columbia  
R.Bartel, B.; Fink, G.R.  
Proc. Natl. Acad. Sci. U.S.A. 91, 6649-53, 1994  
A>Title: Differential regulation of an auxin-producing nitrilase gene family in Arabidop  
A:Reference number: Z24515; NCID:8022831; PMID:8022831  
A:Accession: T52263  
A:Status: preliminary; translated from GB/EMBL/DBD  
A:Molecule type: DNA  
A:Residues: 1-346 <BAR>  
A:Cross-References: UNIPARC:UPI0000001AE2; EMBL:U38845; PIDN:AAB05221.1  
A:Experimental source: cultivar Columbia  
C:Genetics:  
A:Gene: ATSP:T10D17.100; n1c1; N1T1  
A:Map position: 3  
A:Introns: 44/1; 104/1; 202/1; 296/1  
C:Superfamily: nitrilase (carbon-nitrogen hydrolase)  
C:Keywords: hydrolase

Alignment Scores:

Pred. No.:	5,53e-39	Length:	346
Score:	499.50	Matches:	113
Percent Similarity:	54.3%	Conservative:	65
Best Local Similarity:	34.5%	Mismatches:	109
Query Match:	27.4%	Indels:	41
DB:	2	Gaps:	8

US-09-751-299-3 (1-1014) x T49147 (1-346)

QY 13 ATCAAGTCGCGCTGCGGCAAGCGCCGATTCATGATTTGAGGCGATGAC 72  
:::|||||:::|||||:::|||||:::|||||:::|||||  
DB 25 ValArgValThrIleValGlnSerSerThrValYrshnAerPhnProIaThrIleAer 44  
73 AAAACCATGATGATGATGAGAAAGCAGCAGTAATAATGCTCGTGTGATGCCCTTTCCG 132  
:::|||||:::|||||:::|||||:::|||||:::|||||  
DB 45 LysAlaGlnLysYrYrIleValAlaSerLysLysAlaGlnLeuValLeuPhePro 64  
133 GAAACTGGATTCAGGCTACCCCA-----TGGTTCTTTGGCTTGAC 174  
|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
QY 65 GluGluPheIleLysGlyYrProAlaGlyPheArgPheGlyLeuAlaValAlaGlyValHis 84  
175 TAACCAAGCATGGCGAATGCAATTTGTACGCCCAATACCATGAGAACTCATGAGATTGAT 234  
:::|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
DB 85 AsnGluGluGluYrArgAerGluPhe--ArgLysYrThrIleAlaSerAlaIleHisValPro 103  
235 GCGCCTCAAGCTAAAGCGCATTTTCAGATGACGACCGCAAGCGGTTGGGATCATGTCACCTTG 294  
|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
DB 104 GlyProGluValAlaArgLeuAlaAerValAlaArgLysAerHisValTyrLeuValMet 123  
QY 295 GGGATGATGTAACCGGTCGGTGGCGACCCCTTTACATCACTCACTCACTGATGTTACATAGCCATAT 354  
|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
DB 124 GlyAlaIleLeuLysGluGlyYrThrLeuLysCysThrValLeuPheSerProGln 143  
355 GGTGACACCACTTGGGGCGCGCGAAAGTTGAAAGTTGATTTGTTGAAAGTACTTTGTC 414  
|||||:::|||||:::|||||:::|||||:::|||||:::|||||  
DB 144 GlyGlnPheLeuGlyLysHisArgGlyLeuMetProThrIserLeuGlnArgCysIleTrp 163  
415 GCGCAAGGGGATGGTTATGCTAGCGGTTTCCAGACGCTCTGTTGGAAGCTGGGTGGC 474  
|||||:::|||||:::|||||:::|||||:::|||||:::|||||

Db 164 GYINGLYAspGlySerThrIleProValTyrAspThrProIleGlyLysLeuGlyAla 183  
QY 475 TTATCTGTGGAGACACCTTCAACCGGCAAAATACGTTTGATACCAAAATGAA 534  
Db 184 AAlaIleCyStrpGlnAsnArgMetProLeuTyrArgThrAlaLeuTyrAlaYsglyIle 203  
QY 535 GAGATTCATTTGGCG-----GCTTGGCCGAGCTTTAGCTTAT 573  
Db 204 GILuleuTyrCyAlaIleProThrAlaAspGlySerLysGluTrpGlnSerMetLeuHis 223  
QY 574 CCTAATGGCGGCAAGCCCTGGGCGCTGATGCAATGTAACGCGCTTCGAATCTATGCG 633  
Db 224 -----GATATGCTTTGTACA-----GATGACGAAAGCATGCTTGTCTTG 732  
QY 634 GTTGAAGGCAATGCTTCTACTAGCGTGTGCGCTTCGTTCAATCCATGATC--- 690  
Db 226 IIEGlyGlyGlyCyAspValLeuSerAlaCysGlnPheCysGlnArgLysHisPhePro 245  
QY 691 -----GATATGCTTTGTACA-----GATGACGAAAGCATGCTTGTCTTG 732  
Db 246 AspHisProAspTyrLeuPheThrAspTyrLysAspLysGluHisAspSerIleVal 265  
QY 723 GCTGTGTGTGACACTCAAGTATCATAGGCGCTGATGTGTGACTTGTGCGCTCTT 792  
Db 266 SerGlnGlyGly---SerValIleIleSerProLeuGlyGlnValLeuAlaGlyProAsn 284  
QY 793 GCCGAATGAAAGAGGATTTCTCTACGCAACCTTGATCTTGAGTACGATCCTTGCT 852  
Db 285 PheGluSer---GluGlyLeuValThrAlaAspIleAspLeuGlyAspIleAlaArgAla 303  
QY 853 AAAATGGGCGGACCGCTGCTGATTAATCCGCGCCGACATTAAGTCTGCTGTAATA 912  
Db 304 LysLeuTyrPheAspSerValGlyHisTyrSerArgProAspValLeuHisLeuThrVal 323  
QY 913 GATCGACGCTTAATTAACCGGTA 936  
Db 324 AenGluHisProArgLysSerVal 331

## RESULT 9

S22398  
nitrilase (EC 3.5.5.1) - Arabidopsis thaliana  
C.Species: Arabidopsis thaliana (mouse-ear cress)  
C.Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 31-Dec-2004  
C.Accession: S22398  
R.Bartling, D.; Seedorf, M.; Mitchofer, A.; Weiler, E.W.  
Eur. J. Biochem. 205, 417-424, 1992  
A.Title: Cloning and expression of an Arabidopsis nitrilase which can convert indole-3-a  
A.Reference number: S22398; MUID:92209532; PMID:1555601  
A.Accession: S22398  
A.Molecule type: mRNA  
A.Residues: 1-346 <BAR>  
A.Cross-references: UNIPROT:P32961; UNIPARC:UP10000130513; EMBL:X63445; NID:G16399; PIDN  
C.Superfamily: nitrilase (carbon-nitrogen hydrolase)  
C.Keywords: hydrolase

Alignment Scores:  
Pred. No.: 2.05e-38 Length: 346  
Score: 493.50 Matches: 112  
Percent Similarity: 54.3% Conservative: 66  
Best Local Similarity: 34.1% Mismatches: 109  
Query Match: 27.0% Indels: 41  
DB: 2 Gaps: 8

US-09-751-299-3 (1-1014) x S22398 (1-346)

QY 13 ATCAAGTGTGCGCTGTCGACGCGCCGATCTACATGATTTGGAGCGACGGTGC 72  
Db .....  
QY 25 ValAlaGValThrIleValGlnSerThrValTyrAspAspThrProAlaThrIleAsp 44  
Db .....  
QY 73 AAACCATTTAGTTGAGAGAGACGACGCAATTAATGCTCGTGTACGCTTTCCG 132  
Db .....  
45 LysAlaGluLysTyrIleValGlnAlaIleSerLysGlyAlaGluLeuValLeuPhePro 64

QY 133 GAAACTGATTCAGAGTACCA-----TGGTTCTTTGGCTTGAC 174  
Db .....  
QY 65 GYinglyPheIleGlyGlyTyrProArgGlyPheArgPheGlyLeuAlaValGlyAlaHis 84  
Db .....  
QY 175 TCACCAAGATGGGCAATGCAATTTTACGGCCAAATACATGAGACTCATTTGAGT 234  
Db .....  
QY 85 AenGluGlyGlyArgAspGluPhe---ArgLysTyrHisAlaSerAlaIleHisValPro 103  
QY 225 GCGCTCAAGCAAGTACGATTTCAATGACGCAAGCGGTTGGAAATGATGTCACCG 294  
Db .....  
QY 104 GYProGluValAlaArgLeuAlaAspValAlaArgLysValHisValIleValLeuValMet 123  
QY 295 GCGATGATGAAACGGGATGCGTGGACCGCTTACATGATGATGATGATGATGATGAT 354  
Db .....  
QY 124 GYAlaAlaIleGlyGlyGlyGlyTyrThrIleuTyrCysThrValLeuPhePheSerProGln 143  
Db .....  
QY 355 GGTGACACCATTTGGGCGCGCGGCAAGTTGAAACCTTATTTGTTGAAACGTAATTGTC 414  
Db 144 GYinglyPheLeuGlyLysHisIleArgLysLeuMetProThrSerLeuGlyArgCysIleTrp 163  
QY 415 GCGCAAGGAGATGTTTCAATCGTACCGGTTTTCAGAGAGTCTGTTGAAAGCGTGGTGC 474  
Db .....  
QY 164 GYinglyAspGlySerThrIleProValTyrAspThrProIleGlyLysLeuGlyAla 183  
QY 475 TTATCTGTGGAGACACCTTCAACCGGCAAAATACGTTTGATGACCAAAATGAA 534  
Db 184 AAlaIleCyStrpGlnAsnArgMetProLeuTyrArgThrAlaLeuTyrAlaYsglyIle 203  
QY 535 GAGATTCATTTGGCG-----GCTTGGCCGAGCTTTAGCTTAT 573  
Db .....  
QY 574 CCTAATGGCGGCAAGCCCTGGGCGCTGATGCAATGTAACGCGCTTCGAATCTATGCG 633  
Db 224 -----GATATGCTTTGTACA-----GATGACGAAAGCATGCTTGTCTTG 732  
QY 634 GTTGAAGGCAATGCTTCTACTAGCGTGTGCGCTTCGTTCAATCCATGATC--- 690  
Db 226 IIEGlyGlyGlyCyPheValLeuSerAlaCysGlnPheCysGlnArgLysHisPhePro 245  
QY 691 -----GATATGCTTTGTACA-----GATGACGAAAGCATGCTTGTCTTG 732  
Db 246 AspHisProAspTyrLeuPheThrAspTyrLysAspLysGluHisAspSerIleVal 265  
QY 723 GCTGTGTGTGACACTCAAGTATCATAGGCGCTGATGTGTGACTTGTGCGCTCTT 792  
Db 266 SerGlnGlyGly---SerValIleIleSerProLeuGlyGlnValLeuAlaGlyProAsn 284  
QY 793 GCCGAATGAAAGAGGATTTCTCTACGCAACCTTGATCTTGAGTACGATCCTTGCT 852  
Db 285 PheGluSer---GluGlyLeuValThrAlaAspIleAspLeuGlyAspIleAlaArgAla 303  
QY 853 AAAATGGGCGGACCGCTGCTGATTAATCCGCGCCGACATTAAGTCTGCTGTAATA 912  
Db 304 LysLeuTyrPheAspSerValGlyTyrTyrSerArgProAspValLeuHisLeuThrVal 323  
QY 913 GATCGACGCTTAATTAACCGGTA 936  
Db 324 AenGluHisProArgLysSerVal 331

## RESULT 10

S31969  
nitrilase (EC 3.5.5.1) [similarity] - Arabidopsis thaliana  
N.Alternate names: T10D17.90  
C.Species: Arabidopsis thaliana (mouse-ear cress)  
C.Date: 13-Jan-1995 #sequence\_revision 13-Jan-1995 #text\_change 31-Dec-2004  
C.Accession: S31969; T49146; T52258  
R.Bartling, D.; Seedorf, M.; Schmidt, R.C.; Weiler, E.W.  
submitted to the EMBL Data Library, September 1992  
A.Description: Arabidopsis thaliana nitrilases are encoded by a two-member gene family:  
A.Reference number: S31969  
A.Accession: S31969  
A.Molecule type: mRNA



[illegible]

```

RESULT 12
T49148
nitrilase (EC 3.5.5.1) 3 [imported] - Arabidopsis thaliana
N:Alternate names: protein T10D17.110
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 02-Jun-2000 #sequence revision 02-Jun-2000 #text_change 31-Dec-2004
C:Accession: T49148; T52261; T52264
R:D'Angelo, M.; Vezzi, A.; Modesto, D.; Pigazzi, M.; Valle, G.; Mewes, H.W.; Rudd, S.; I
submitted to the Protein Sequence Database, April 2000
A:Reference number: Z5017
A:Accession: T49148
A:Status: preliminary
A:Molecule type: DNA

```

A.Residues: 1346 <DAN>  
A.Cross-references: UNIPROT:P46010; UNIPARC:UPI000000BDB6; EMBL:ALJ35865; GSPDB:GN00061.  
A.Experimental source: cultivar Columbia; BAC clone T10D17  
R.Hillebrand, H.; Bartling, D.; Weiler, E.W.  
Plant Mol. Biol. 36, 89-99, 1998  
A.Title: Structural analysis of the ntc2/ntc1/ntc3 gene cluster encoding nitrilases, enz  
A.Reference number: Z26007; MUID:98145459; PMID:9464465  
A.Accession: T52261  
A.Status: preliminary; translated from GB/EMBL/DDBJ  
A.Molecule type: DNA  
A.Residues: 1-346 <HIL>  
A.Cross-references: UNIPARC:UPI000000BDB6; EMBL:Y07648; PIDN:CAA68936.2  
A.Experimental source: cultivar Columbia  
R.Bartel, B.; Fink, G.R.  
Proc. Natl. Acad. Sci. U.S.A. 91, 6649-53, 1994  
A.Title: Differential regulation of an auxin-producing nitrilase gene family in Arabidops  
A.Reference number: Z24515; MUID:8022831; PMID:8022831  
A.Accession: T52264  
A.Status: preliminary; translated from GB/EMBL/DDBJ  
A.Molecule type: mRNA  
A.Residues: 1-346 <BAR>  
A.Cross-references: UNIPARC:UPI000000BDB6; EMBL:U09959; PIDN:AAA19627.1  
A.Experimental source: cultivar Columbia  
C.Genetics:  
A.Gene: ATSP:T10D17.110; ntc3  
A.Map position: 3  
A.Introns: 44/1, 104/1, 202/1, 296/1  
C.Superfamily: nitrilase (carbon-nitrogen hydrolase)  
C.Keywords: hydrolase

Alignment Scores:

Pred. No.:	3,95e-38	Length:	346
Score:	490.50	Matches:	115
Percent Similarity:	55.9%	Conservative:	65
Best Local Similarity:	35.7%	Mismatches:	113
Query Match:	26.9%	Indels:	29
B:	2	Gaps:	8

US-09-751-299-3 (1-1014) x T49148 (1-346)

OY 13 ATCAAGTCGCTGCCTGTGAAGCGCCCATCTTACATGATTGGAGCGACGCTGAC 72  
:::|||||  
Db 25 VALAGVAlThrIleValInserSerThrValTyraSnSptPrProAlatThrLeuasp 44  
|||  
OY 73 AAAACATTGACTTGATGAGAAGACGACGAAATTAATCTGCTGCATGCGCTTCCG 132  
|||  
Db 45 LysAlaGlulysPheIleValGlulAlaAserylsglyAlalysLeuValLeuphePro 64  
|||  
OY 133 GAACCTGGATTCCAAGCTTACCA-----TGTTCTTTGGCTTAC 174  
|||  
Db 65 GlulAlaPheIleGlyTyrProArgGlyPheArgPheGlyLeuAlaValAlHis 84  
|||  
OY 175 TCACCAGCATGGGCAATGCAATTTGTACGCCCAATACCATGAGAACTAATGAGTGGAT 234  
:::  
Db 85 AengluGlulYArGaspGluPhe--ArgasnTyChisAlaSerAlaIleValPro 103  
|||  
OY 235 GGCCCTCAAGCTAAGGGCATTTTCAGATGACAAGCGCGTGGGAATCATGTCACCCCTG 294  
|||||  
Db 104 GLYPFGIuValGlulurGluAlaGlulAlaGlyLysAsnbnValHisLeuValMet 123  
|||  
OY 295 GGGATGATGAGAACGGGTCGTTGGCACCTTTACATGATGATGAGTTCATATGAGCATAT 354  
|||  
Db 124 GlyAlaIleGlyLysAspGlyTyrThrLeuTyrCythrThrAlaLeuPhePheSerProIn 143  
|||  
OY 355 GGTGACACCAATTTGGGGCCCGCGAAGATTGAAACTTACTTTTGTGAACGTACTTTGTT 414  
|||  
Db 144 GlyInPheLeuGlyLysHisArgLysValMetProThrSerLeuGluArgCylIenTrp 163  
|||  
OY 415 GGCAGAGGGAGGTTCATCGCTAGCGGTTTTCAGACGTCGTGTGGAAGCGTGGCGC 474  
|||  
Db 164 GlyIngLyAspGlySerThrIleProValTyrAspThrProIleGlyLysIleGlyAla 183  
|||  
OY 475 TTATGCTTTGGACACCTTCAACCGCTTACCAAAATAACCTTTGTATGCAAAATGAA 534





```
Db      171 IleIleTrpGlyPheGlyAspGlySerThrIleProValIlyrAspThrProLeuGlyLys 190
QY      466 CTGGGTGGCTTATGCTGTGGAGACACCTTCAACCGCTAACAAAATAGCCTTGTATGCA 525
        ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      191 IleGlyAlaLeuIleCysTrpGlyAsnLysMetProLeuLeuArgThrAlaLeuIlyrGly 210
QY      526 CAAAATGAAGAGATTCAATTGTGGCGCTTGGCCGAGCTTTAGCCTTTATCCTAATGCCGCG 585
        ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      211 LysGlyIleGlyIleIlyrCysAla-----ProThrAlaAsp 222
QY      586 AAGCCCTGGGGCCTGATGTCATATGTAAGCGCCTTCGAATCTATGCCGTTGAAGGCAA 645
        ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      223 SerArg-----GlnValTrpGlnAlaSerMetThrHisIleAlaLeuGlyGly 239
QY      646 TGCTTCGTAAGCGTCGTGTCGCTCGTTTCACAAATCC----- 684
        ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      240 CysPheValLeuSerAlaAsnGlnPheCysArgArgLysAspIlyrProProProGlu 259
QY      685 ---ATGATCGATATGCTTTGTACAGATGACGAAAAGCATGCGCTTCTGCTGTGTGT 741
        ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      260 TyrValPheThrGlyLeuGlyGluGluProSerProAspThrValValCysProGlyGly 279
QY      742 GGACACTCAGCTATCATAGGCGCTGATGTGTGTGATGACTTGTGCGGCTTGTGCCGAAAT 801
        ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      280 -----SerValIleIleSerProSerGlyGluValLeuAlaGlyProAsnIlyrGly 297
QY      802 GAAGAGGATATTCTCTACGCAACCTTGATCTGGAGTACGCATCTTGTAAATGCGG 861
        ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      298 ---GlnAlaLeuIleThrAlaAspLeuAspLeuGlyGluIleValArgAlaLysPheAsp 316
QY      862 GCAGACCTGCTGTGTCATTATTCGCCGTCGACATTACTCGCTGCTAATAGATCGCAGC 921
        ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      317 PheAspValValGlyHisIlyrAlaArgProGluValIleuSerLeuValValAsnAspGln 336
QY      922 CCTAATTACCGGTA 936
        ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      337 ProHisLeuProVal 341
```

Search completed: April 27, 2006, 01:38:10  
Job time : 45.981 secs



GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: April 27, 2006, 01:07:06 ; Search time 49.8958 Seconds  
(without alignments)  
2868.171 Million cell updates/sec

Title: US-09-751-299-3  
Perfect score: 1826  
Sequence: 1 atgaagaagcatcatcaagt.....cgcaactcgagaatttga 1014

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2166443 segs, 705528306 residues  
Total number of hits satisfying chosen parameters: 4332886

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:  
-MODE=frame+ n2p.model  
-Q=/abs/ABSSWEB.spool/US09751299/runat\_26042006\_090034\_18631/app\_query.fasta\_1  
-DB=UniProt -QFMT=fastan -SUFFIX=rup -MINMATCH=0.1 -LOOPEXT=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human4.cdi -LIST=45  
-DOCALLIGN=200 -THR SCORE=pct -THR MAX=100 -THR MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pro -NORM=ext -HEADSITE=500 -MINLEN=0 -MAXLEN=200000000 -HOST=abss802p  
-USER=US09751299 @CN 1 1.580 @runat\_26042006\_090034\_18631 -NCPU=6 -ICPU=3  
-NO\_MMAP -NEG SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG -DEV TIMEOUT=120  
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-XGAPOP=10 -XGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : UniProt\_05.80.\*  
1: uniprot\_sprot.\*  
2: uniprot\_trembl.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1765	96.7	337	2	Q6RWE0_9222Z
2	1343	73.5	337	2	Q6RWO3_9222Z
3	1343	73.5	337	2	Q6RWO3_9222Z
4	1187	65.0	336	2	Q500U1_PSSBY
5	1139	64.8	347	2	Q88B32_PSSSM
6	1139	62.4	333	2	Q6RWO3_9222Z
7	1136	62.4	332	2	Q6RWO3_9222Z
8	1081	59.2	345	2	Q6RWS5_9222Z
9	1051	57.6	334	2	Q6RWS5_9222Z
10	1042	57.1	353	2	Q6RWS2_9222Z
11	1039	56.9	348	2	Q6RWS2_9222Z
12	1037	56.8	336	2	Q6RWF6_9222Z
13	1037	56.8	333	2	Q6RWF0_9222Z
14	1033	56.6	337	2	Q6RWS3_9222Z
15	1033	56.6	338	2	Q6RWB4_9222Z
16	1028.5	56.3	358	2	Q6RW10_9222Z

17	1022.5	56.0	337	2	Q6RWN8_9222Z	Q6RWN8 unclutred
18	1021.5	55.9	337	2	Q6RWN9_9222Z	Q6RWN9 unclutred
19	1020.5	55.9	384	2	Q6RWK6_9222Z	Q6RWK6 unclutred
20	1019.5	55.8	338	2	Q6RWG6_9222Z	Q6RWG6 unclutred
21	1018	55.8	338	2	Q6RWM7_9222Z	Q6RWM7 unclutred
22	1008	55.2	337	2	Q6RWM7_9222Z	Q6RWM7 unclutred
23	1008	55.2	346	2	Q6RWS5_9222Z	Q6RWS5 unclutred
24	999.5	54.7	338	2	Q6RWS2_9222Z	Q6RWS2 unclutred
25	999	54.7	338	2	Q6RWR6_9222Z	Q6RWR6 unclutred
26	997.5	54.6	336	2	Q5YUW5_NOCFA	Q5YUW5 nocardia fa
27	994.5	54.5	338	2	Q6RWO4_9222Z	Q6RWO4 unclutred
28	993	54.4	337	2	Q6RWH1_9222Z	Q6RWH1 unclutred
29	980.5	53.7	327	2	Q6RWH9_9222Z	Q6RWH9 unclutred
30	976	53.5	352	2	Q6RWH0_9222Z	Q6RWH0 unclutred
31	973.5	53.3	336	2	Q6RWS5_9222Z	Q6RWS5 unclutred
32	971.5	53.2	341	2	Q6RWR4_9222Z	Q6RWR4 unclutred
33	937.5	51.3	335	2	Q6RWF7_9222Z	Q6RWF7 unclutred
34	936	51.3	331	2	Q70608_PSSPU	Q70608 pseudomonas
35	929.5	50.9	337	2	Q6RWM3_9222Z	Q6RWM3 unclutred
36	924	50.6	350	2	Q5EG61_PSSFL	Q5EG61 pseudomonas
37	919.5	50.4	334	2	Q6RWS9_9222Z	Q6RWS9 unclutred
38	918.5	50.3	334	2	Q89GE3_BRADJA	Q89GE3 bradyrhizob
39	914.5	50.1	337	2	Q6RWM9_9222Z	Q6RWM9 unclutred
40	906	49.6	345	2	Q6RWN2_9222Z	Q6RWN2 unclutred
41	898.5	49.2	349	2	Q6RWS0_9222Z	Q6RWS0 unclutred
42	881	48.2	346	2	Q6RWM1_9222Z	Q6RWM1 unclutred
43	836.5	45.8	365	1	NRL2_RHORA	Q03217 rhodococcus
44	819	44.9	356	1	NRLA_ALCFA	P20960 alcaligenes
45	800	43.8	383	1	NRL1_RHORA	Q02068 rhodococcus

## ALIGNMENTS

RESULT 1  
ID Q6RWE0\_9222Z PRELIMINARY; PRT; 337 AA.  
AC Q6RWE0;  
DT 05-JUL-2004 (TREMBLrel. 27, Created)  
DT 05-JUL-2004 (TREMBLrel. 27, Last sequence update)  
DT 05-JUL-2004 (TREMBLrel. 27, Last annotation update)  
DE Nitrilase (EC 3.5.5.7).  
GN ORFName=BD5220;  
DE Nitrilase (EC 3.5.5.7).  
OS unclutred organism.  
OC unclassified; environmental samples.  
OX NCBI\_TaxID=155900;  
RN [1]  
RP NUCLEOTIDE SEQUENCE.  
RX PubMed=15066841; DOI=10.1128/AEM.70.4.2429-2436.2004;  
RA Robertson D.E., Chaplin J.A., Desantis G., Podar M., Madden M.,  
RA Chi E., Richardson T., Milan A., Miller M., Weiner D.P., Wong K.,  
RA McQuaid J., Farwell B., Preston L.A., Tan X., Sneed M.A., Keller M.,  
RA Mathur E., Kretz P.L., Burk M.J., Short J.M.;  
RT "Exploring nitrilase sequence space for enantioselective catalysis.";  
RL Appl. Environ. Microbiol. 70:2429-2436(2004).  
DR EMBL: AY487562; AAR97509.1; -; Genomic DNA.  
DR GO: GO:0018762; F:aliphatic nitrilase activity; IEA.  
DR GO: GO:0016810; F:hydrolase activity; acting on carbon-nitrog. . .; IEA.  
DR InterPro: IPR00132; N/CN hydrolase.  
DR InterPro: IPR003010; N/CN hydrolase.  
DR Pfam: PF00795; CN hydrolase; 1.  
DR PROSITE: PS0263; CN HYDROLASE; 1.  
DR PROSITE: PS00921; NITRIL\_CHT\_2; 1.  
KW Hydrolase.  
SQ SEQUENCE 337 AA; 36783 MW; 25D2DE1B820983B2 CRC64;  
Alignment Scores:  
Pred. No.: 3.08e-149  
Score: 1765.00  
Percent Similarity: 100.0%  
Best Local Similarity: 100.0%  
Query Match: 96.7%  
Length: 337  
Matches: 337  
Conservative: 0  
Mismatch: 0  
Indels: 0  
Gaps: 0

US-09-751-299-3 (1-1014) x Q6RWO3\_9ZZZZ (1-337)

1 ATGAAAGAGCTATCAAGTGGCTGGTGCAGCGCCCGCATCTACATGATTTGGAG 60  
 1 Metlysegluaalieleysvalalacysvalalgnalalaprolietyleymelaaspleuglu 20  
 61 GCGACGGTGACAAACCATTTGAGTTGATGGAAGAGGACGATTAATGCTGCTGTG 120  
 21 Alathrvalaspylthrlilegluleumetgluualalalargasnaalalargleu 40  
 121 ATGCGCTTTCCGGAACCTTTGATTCAGGCTTACCAATGGTTCTTTGGCTTGACACCA 180  
 41 lleaiaheplogluthrtripleprogllytrprottrpneutrleupneaspserpro 60  
 181 GCATGGGGAATGCAATTTGTACGCAATACCAATGAACTCATTTGAGAGTTGAGGCCCT 240  
 61 Alatrpalameglmphevalargglintyrhisgluanserleugluueuaspglypro 80  
 241 CAAGCTAAGCGCATTTCAAGATSCAGCCAAAGCGTTGGGATCATGCTCCTCGGGGATG 300  
 81 Glnalalyasargileaserasplalalyasargleuglyilemetvalthrleuuglymet 100  
 301 AGTGAACGGGTGGTGGACCCCTTTACATCAGTGGTGTGATAGCGCATATGCTGAC 360  
 101 Sergluargvalaglylthreuleuyrileserglnttrpneillelyasphanuglyasp 120  
 361 ACCATTGGGGCGCGGGAAGTTGAACCTACTTTTGTGAACGTACTTTGTTGGCGGA 420  
 121 Thrilleglyalalargarglyseuleuyrprothrphelvalgluhgrthrleupheuglylu 140  
 421 GGGGATGTTTCATCGGTAGCGGTTTTCAGACGCTCTGTTGGAAGGCTGGTGGCTTATGC 480  
 141 Glyaspglyseraserleuvalalphegluthrsevalglyargleuglyyleuqys 160  
 481 TGTGGGAGACCTTCAACCGCTAACAATAAGCTTTGATGCACAAATGAAGAGATT 540  
 161 Cyslrpoglunhlsleuonprouleumthlyrvalaleuuyrhalaglnanugluuile 180  
 541 CATTTGCGGCTTGGCGAGCTTTAGCCTTTATCTTAATGCGGGAAGCCCTGGGGCCT 600  
 181 Hiecyahlaalatrproserpneaserleuuyrproasnaalalalyaslaaleuglypro 200  
 601 GATGTCATGATGAGCGGCTCTCGAATCTAGCGCTTGAAGGCAATGCTTGATCTAGCG 660  
 201 Aspvalasnaalalalaserargiletyralavalagluglylncysphevalleualla 220  
 661 TCGGTGCGGCGCTTTCACATCGATCGATGCTTTGTGTACATGACGAAAGCAT 720  
 221 Serasyalaleuvalaserglinsermetileasmetleuuyrthrapsaspglulysheis 240  
 721 GCGTGTCTTTCGAGTGTGTGTGACACTCAGCTATCATAGGGCTGATGTGTGTGACTTG 780  
 241 Alaleuueuleuvalaglylgllyhisaserargilelleuglyproaspglyglyasprou 260  
 781 GTGCGGCTCTTTCGGAATAAGAGAGGTATTCCTACGAAACCTTGATCTCGAGTA 840  
 261 Valalaproleuvalagluansglunuglylileuuyrhalasneueaspprogllyal 280  
 841 CGCATCTTGTGTAATAAGCGGACAGACCTGCTGTCTATTTCCGCTCCGACATTAAT 900  
 281 Argilleuualalyasmechlaalapsproliaglyhisetyrserargproaspllethr 300  
 901 CGCTTGATTAATGATCGAGCCCTTAATTAACCGGTAGTTGAATTTGAAGGTGATCTTGT 960  
 301 Argleuleuulleasparaserprolyseuprovalalagluillegluglyasprouarg 320  
 961 CCTTACGCTTTGGTGAAGCGCTGAGACGGGTGGGCACTGAGAAATTT 1011  
 321 Protyralaleuuglylysalaserglunhrglyalaglnleuugluuile 337

RESULT 2  
 Q6RWO3\_9ZZZZ

ID Q6RWO3\_9ZZZZ PRELIMINARY; PRT; 337 AA.  
 AC Q6RWO3;  
 DT 05-JUL-2004 (TREMBLREL. 27, Created)  
 DT 05-JUL-2004 (TREMBLREL. 27, Last sequence update)  
 DT 05-JUL-2004 (TREMBLREL. 27, Last annotation update)  
 DE Nucleic acid  
 GN Nucleic acid (EC 3.5.5.7).  
 OS unclassified; environmental samples.  
 OC unclassified; environmental samples.  
 CX NCBI\_TaxId=155900;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RA PubMed=15066841; DOI=10.1128/AEM.70.4.2429-2436.2004;  
 RA Robertson D.E., Chaplin J.A., Desantis G., Podar M., Madden M.,  
 RA Chi E., Richardson T., Milan A., Miller M., Weiner D.P., Wong K.,  
 RA McQuaid J., Farwell B., Preston L.A., Tan X., Sneed M.A., Keller M.,  
 RA Machur E., Kretz P.L., Burk M.J., Short J.M.,  
 RT "Exploring nitrite sequence space for enantioselective catalysis."  
 RL Appl. Environ. Microbiol. 70:2429-2436(2004).  
 DR EMBL; AY487449; AAR97396.1; -; Genomic DNA.  
 DR GO; GO:0018762; F:allophatic nitritease activity; IEA.  
 DR GO; GO:0016810; F:hydrolyase activity; acting on carbon-nitrog. . .; IEA.  
 DR GO; GO:0006807; P:nitrogen compound metabolism; IEA.  
 DR InterPro; IPR000132; N/CN hydrolase.  
 DR InterPro; IPR003010; Nucleic acid hydrolase.  
 DR Pfam; PF00795; CN hydrolase; 1.  
 DR PROSITE; PS50263; CN HYDROLASE; 1.  
 DR PROSITE; PS00920; NITRIL CHT 1; 1.  
 DR PROSITE; PS00921; NITRIL\_CHT\_2; 1.  
 KW Hydrolyase. 337 AA; 36697 MW; 940993D5AB69582E CRC64;  
 SQ SEQUENCE

Alignment Scores:  
 Pred. No.: 2,18e-111 Length: 337  
 Score: 1343.00 Matches: 253  
 Percent Similarity: 84.2% Conservative: 29  
 Best Local Similarity: 75.5% Mismatches: 47  
 Query Match: 73.5% Indels: 6  
 DB: 2 Gaps: 1

US-09-751-299-3 (1-1014) x Q6RWO3\_9ZZZZ (1-337)

1 ATGAAAGAGCTATCAAGTGGCTGGTGCAGCGCCCGCATCTACATGATTTGGAG 60  
 1 Metlysegluaalieleysvalalacysvalalgnalalaprovalpheuaspserpro 20  
 61 GCGACGGTGACAAACCATTTGAGTTGATGGAAGAGGACGATTAATGCTGCTGTG 120  
 21 Alathrvalaspylthrlilegluleuullegluualalalargasnaalalargleu 40  
 121 ATGCGCTTTCCGGAACCTTTGATTCAGGCTTACCAATGGTTCTTTGGCTTGACACCA 180  
 41 lleaiaheplogluthrtripleprogllytrprottrpneutrleupneaspserpro 60  
 181 GCATGGGGAATGCAATTTGTACGCAATACCAATGAACTCATTTGAGAGTTGAGGCCCT 240  
 61 Alatrpalameglmphevalargglintyrhisgluanserleugluueuaspglypro 80  
 241 CAAGCTAAGCGCATTTCAAGATSCAGCCAAAGCGTTGGGATCATGCTCCTCGGGGATG 300  
 81 Glnalalyasargileaserglunhrglyalaglnarglyllyrvalalaleuglytyr 100  
 301 AGTGAACGGGTGGTGGACCCCTTTACATCAGTGGTGTGATAGCGCATATGCTGAC 360  
 101 Sergluargvalaglylthreuleuyrileserglnttrpneillelyasphanuglyasp 120  
 361 ACCATTGGGGCGCGGGAAGTTGAACCTACTTTTGTGAACGTACTTTGTTGGCGGA 420  
 121 Thrllaglyleuvalargarglyseuleuyrprothrphelvalgluhgrthrleupheuglylu 140  
 421 GGGGATGTTTCATCGGTAGCGGTTTTCAGACGCTCTGTTGGAAGGCTGGTGGCTTATGC 480  
 141 Glyaspglyseraserleuvalalphegluthrsevalglyargleuglyyleuqys 160

QY 481 TGTGGAGACCTTCAACCGCTAACAAAATACGTTTGTATGACAAAATGAGAGATT 540  
 |||||  
 Db 161 CysTTPGIuHleuInProleuSerLyTyraLeuTyraLaInaInguInuLe 180  
 541 CATTGTGGCGCTTGGCGGAGCTTTAGCCTTTATCTTAATGCGCGAAGCCCTGGGCGCT 600  
 |||||  
 Db 181 HispHeAlaAlaIatProSerPheSerIleTyraArgInAlaInthGluValleuGlyPro 200  
 601 GATGTCAATGTAGCGGCGCTCTGAAATCTATGCCGTGGAAGGGAATGCTTGTGATAGCG 660  
 : : : : :  
 Db 201 GluValaInValaAlaIaSerArgIleTyraIaValaGluGlyInCySpheValleuAla 220  
 661 TCGTGTGGCGCTCTTGCATCAATCATCATGATATGCTTTGTACAGATCGAAGAAAGCAT 720  
 |||||  
 Db 221 SerCyAlaLeuValaSerProGluMetIleGluMetLeuCyThrAspGluSerLyHis 240  
 721 GCGTGTCTTCTGCGTGTGGTGAGACACTACGATATAGAGGCGTGTATGATGCTTGTGACTTG 780  
 : : : : :  
 Db 241 SerLeuLeuGlnAlaIaGlyLyTyraSerArgIleIleGlyProAspGlySerAspLeu 260  
 781 GTGCGGCGCTTGTGCGGAAATGAAAGGATTTCTTACGAAACCTTGATCCTGGAGTA 840  
 |||||  
 Db 261 AlaArgProLeuGlyGluInaInguInuGlyIleLeuTyraIaThrLeuAspProAlaAla 280  
 841 CGCATCTTGTAAATGCGGAGACCGCTGCTGCTATTTCCGTCGCGACATTAAT 900  
 |||||  
 Db 281 ArgIleTyraIaLySerThrAlaIaAspProAlaGlyHisTyraSerArgProAspValThr 300  
 901 CGCTTGCTAATAGATCGACGCGCTTAATTAATCCGCTAGTGAATGGAAGTATCTTCGT 960  
 |||||  
 Db 301 ArgLeuLeuIleAsnArgSerAlaInaInProValaGluValaGlyArgGluIle--- 319  
 961 CCTTACGCTTTGGGTAAAGCGTCTGAGACGGGTGCGCAACTCGAA 1005  
 |||||  
 Db 320 -----ProAlaSerAlaGlnGlyPheGluValaGlu 329

## RESULT 3

Q6RM07\_92222

Q6RM07\_92222 PRELIMINARY; PRT; 337 AA.

AC 06RM07;  
 05-JUL-2004 (TRMBLrel. 27, Created)  
 DT 05-JUL-2004 (TRMBLrel. 27, Last sequence update)  
 DT 05-JUL-2004 (TRMBLrel. 27, Last annotation update)  
 DE Nitrilase (EC 3.5.5.7).  
 GN ORFNames=BD5274;  
 OS unclassified organism.  
 OC unclassified; environmental samples.  
 OX NCBI\_TaxID=155900;  
 RN [1]

RP NUCLEOTIDE SEQUENCE.  
 RX PubMed=15066841; DOI=10.1128/AEM.70.4.2429-2436.2004;  
 RA Robertson D.E., Chaplin J.A., Desantis G., Fodor M., Madden M.,  
 RA Chi E., Richardson T., Milan A., Miller M., Weiner D.P., Wong K.,  
 RA McQuaid J., Farwell B., Preston L.A., Tan X., Sneed M.A., Keller M.,  
 RA Machur E., Kretz P.L., Burk M.J., Short J.M.,  
 RT "Exploiting nitrilase sequence space for enantioselective catalysis.";  
 RL Appl. Environ. Microbiol. 70:2429-2436 (2004).  
 DR EMBL; AY487445; AAC97392.1; -; Genomic DNA.  
 DR GO; GO:0018762; F:allophatic nitrilase activity; IEA.  
 DR GO; GO:0016810; F:hydrolyase activity; acting on carbon-nitrog. . .; IEA.  
 DR GO; GO:0006807; P:nitrogen compound metabolism; IEA.  
 DR InterPro; IPR000132; N/CN hydrolase.  
 DR InterPro; IPR003010; Ntase/CNhydase.  
 DR Pfam; PF00795; CN hydrolase; 1.  
 DR PROSITE; PS50263; CN HYDROLASE; 1.  
 DR PROSITE; PS00920; NITRIL\_CHT 1; 1.  
 DR PROSITE; PS00921; NITRIL\_CHT\_2; 1.  
 KW Hydrolase. 337 AA; 36769 MW; 940905AB69582E CRC64;

Alignment Scores: 2.18e-111 Length: 337  
 Pred. No.:

Score: 1343.00 Matches: 253  
 Percent Similarity: 84.2%  
 Best Local Similarity: 75.5%  
 Query Match: 73.5%  
 DB: 2 Gaps: 1

US-09-751-299-3 (1-1014) x Q6RM07\_92222 (1-337)

QY 1 ATGAAAGAGCTATCAAGTCCCTGCTGCAAGCCGCCGATCTACATGATTTGGAG 60  
 |||||  
 Db 1 MetLySGluAlaIleLyValaAlaCyValaGlnAlaIaProValaPheLeuAspLeuAsp 20  
 61 GCGACGGTGACAAACCATGATGATGAAAGAGACGCTAATTAATGCTGTGCG 120  
 |||||  
 Db 21 AlaThrValaAspLyThrValaAlaLeuIleGluGlnAlaAlaArgAsnGlyAlaArgLeu 40  
 121 ATGCGCTTCCGGAACCTTGATATTCAGAGTCAACCATGATGATTTCTTTGGCTTGACTACCA 180  
 |||||  
 Db 41 IleAlaPheProGluThrTriPleProGlyTyraProIlePheLeuTriPleuAspSerPro 60  
 181 GCATGGGCAATGCAATTTGTACGCCAATACATGAACTCATTTGAGTGGAGCCCT 240  
 |||||  
 Db 61 AlaTrpGlyMetGlnPheValaArgTyraHisGluAsnSerLeuValaLeuAspSerPro 80  
 241 CAAGCTAAGCGCATTTTCAGATGCGACGCGGCTTGGGAATCATGATGATCCTCGGGGATG 300  
 |||||  
 Db 81 GlnAlaLyAspArgIleSerGluAlaIaGlnAlaArgAlaGlyIleTyraValaAlaLeuGlyTy 100  
 301 AGTGAACGGGTGCGTGACACCTTTACATGATGCTGATGCTTATGAGGATATGATGATGAC 360  
 |||||  
 Db 101 SerGluAspValaSerGlyThrLeuTyraMetGlyGlnTriPleuIleAspAspLySerGly 120  
 361 ACCATTGGGGCCCGCGCAAGTTGAACCTTACTTTGTGAACGATCTTTGTTCGGGCA 420  
 |||||  
 Db 121 ThrAlaGlyLeuAspArgLyLeuAspProThrHisValaGluArgThrLeuPheGlyGlu 140  
 421 GGGGATGGTTCATGCTGACCGGTTTTCAGACGCTGTGTGAAGGCTGGGTGGCTTATGAC 480  
 |||||  
 Db 141 GlyAspGlySerSerLeuSerThrPheAspThrProLeuGlyValaLeuGlyGlyLeuCyS 160  
 481 TGTGGGAGACCTTCAACCGCTTAACAAATACGCTTGTATGACAAATGAAGAGATT 540  
 |||||  
 Db 161 CysTrpGluHleuGlnProleuSerLyTyraLeuTyraIaGlnAlaInaInguIle 180  
 541 CATTGTGGCGCTTGGCGGAGCTTTAGCCTTTATCTTAATGCGCGCAAGCCCTGGGCGCT 600  
 |||||  
 Db 181 HispHeAlaAlaIatProSerPheSerIleTyraArgInAlaInthGluValleuGlyPro 200  
 601 GATGTCAATGTAGGCGCTCTCGAATCTATGCTGTAAGGCGCAATGCTTGTACTACGC 660  
 : : : : :  
 Db 201 GluValaInValaAlaIaSerArgIleTyraIaValaGluGlyGlnCySpheValleuAla 220  
 661 TCGTGTGGCGCTTTCACATCATCATGATATGCTTTGTACAGATGAGCAAAAGCAT 720  
 : : : : :  
 Db 221 SerCyAlaLeuValaSerProGluMetIleGluMetLeuCyThrAspGluSerLyHis 240  
 721 GCGTGTCTTCTGCGTGTGGTGAGACACTACGATATCATAGGCGCTGATGTGTGACTTG 780  
 : : : : :  
 Db 241 SerLeuLeuGlnAlaIaGlyLyTyraSerArgIleIleGlyProAspGlySerAspLeu 260  
 781 GTGCGGCGCTTGTGCGGAAATGAAAGGATTTCTTACGAAACCTTGATCCTGGAGTA 840  
 |||||  
 Db 261 AlaArgProLeuGlyGluInaInguInuGlyIleLeuTyraIaThrLeuAspProAlaAla 280  
 841 CGCATCTTGTAAATGCGGAGACCGCTGCTGCTATTTCCGTCGCGACATTAAT 900  
 |||||  
 Db 281 ArgIleTyraIaLySerThrAlaIaAspProAlaGlyHisTyraSerArgProAspValThr 300  
 901 CGCTTGCTAATAGATCGACGCGCTTAATTAATCCGCTAGTGAATGGAAGTATCTTCGT 960  
 |||||  
 Db 301 ArgLeuLeuIleAsnArgSerAlaInaInProValaGluValaGlyArgGluIle--- 319  
 961 CCTTACGCTTTGGGTAAAGCGTCTGAGACGGGTGCGCAACTCGAA 1005



DR TIGR; PSPT00189; -  
 DR GO; GO:0016810; F:hydrolyase activity, acting on carbon-nitrog. . .; IEA.  
 DR GO; GO:0006807; P:nitrogen compound metabolism; IEA.  
 DR InterPro; IPR000132; N/CN hydrolase.  
 DR InterPro; IPR003010; Ntase/CNhydase.  
 DR Pfam; PF00795; CN hydrolase; 1.  
 DR PROSITE; PS50263; CN HYDROLASE; 1.  
 DR PROSITE; PS00920; NITRIL\_CHT\_1; 1.  
 DR PROSITE; PS00921; NITRIL\_CHT\_2; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 347 AA; 37564 MW; B671F783D7411308 CRC64;

Alignment Scores:

Pred. No.:	4e-97	Length:	347
Score:	1184.00	Matches:	218
Percent Similarity:	80.4%	Conservative:	37
Best Local Similarity:	68.8%	Mismatches:	62
Query Match:	64.8%	Indels:	0
DB:	2	Gaps:	0

US-09-751-299-3 (1-1014) x Q88B32\_PSESM (1-347)

```

QY 1 ATGAAAGAGCTATCAAGTCCCTGCGTGAAGCCCGCATCTACATGATTGGAG 60
   |||||
Db 12 MetlysaLProleuylValAlaCysValGlnAlaIleProMetPheLeuAsnLeuasp 31
   |||||

QY 61 GCGACGGTGGACAAACCTTGAATTGATGGAAGACGACGCTAATATGCTGCTG 120
   |||||
Db 32 AlatrValaAspLysThrIleAlaLeuIleGlnAlaIleAlaIleValaIleGlyLeu 51
   |||||

QY 121 ATGCGCTTCCGGAACCTTGATTCAGGCTACCCATGGTCTTCTGGCTGATCACA 180
   |||||
Db 52 IlealapeProGlnThrIleProGlyIleProTyrPheLeuThrPheLeuAspIlePro 71
   |||||

QY 181 GCATGGGCAATGCAATTTGTAGCCCAATACCATGAGAACTCATGAGTTGAGTGCCT 240
   |||||
Db 72 AlatrPasnMetProLeuValHisArgTyrHisGlnGlnSerLeuValLeuAspSerAla 91
   |||||

QY 241 CAAGTAAAGCCGATTTCAATGATGACACCGCGGTTGGGAATATGTCACCTGGGAGATG 300
   |||||
Db 92 GlnAlaAspArgIleSerAspAlaIleArgGlnHisGlyIleTyrValValLeuGlyTyr 111
   |||||

QY 301 AGTGAACGGGTCCGTGGACACCTTACATGAGTGGTTCATGAGGAGTAATGATGAC 360
   |||||
Db 112 SerGlnArgSerLysAlaSerLeuTyrIleGlyGlnThrIleIleAspSerGlnGlyLeu 131
   |||||

QY 361 ACCATTGGGGCCCGCGAAGTTGAACCTACTTTGTTGAACGTACTTGTTCGCGCAA 420
   |||||
Db 132 ThrValGlyThrArgArgLysLeuLysAlaThrHisValGlnArgThrLeuPheGlyLeu 151
   |||||

QY 421 GGGGATGGTTCATCGCTAGCGGTTTTCAGACGCTGTTGGAAAGCTGGGCTTATGC 480
   |||||
Db 152 GlyAspGlyAlaSerLeuArgProPheGlnThrProValGlyValLeuGlyAlaLeuGly 171
   |||||

QY 481 TGTGGGAGGACCTTCAACCGCTACAAATACGTTGTTGTTGCAAAATGAAGATT 540
   |||||
Db 172 CysTrpGlnHisSerGlnProLeuSerLysTyrAlaMetTyrAlaGlnAsnIleGlnIle 191
   |||||

QY 541 CATGTGGCGGTGGCCGAGCTTTAGCCTTATCTTAATGCGCGAAGCCCTGGGGCT 600
   |||||
Db 192 HisValAlaIleAlaIleProSerPheSerLeuTyrGlnAsnAlaThrArgAlaLeuGlyPro 211
   |||||

QY 601 GATGTCATGTAGCGGCTCTGCAATCTATGCGCTTGAAGGGCAATGCTGCTACTACG 660
   |||||
Db 212 GlnValaLeuThrAlaIleSerArgValTyrAlaIleGlnGlyGlnCysPheValIleAla 231
   |||||

QY 661 TCGTGGCGCTCGTTCAATTCATGATCGATATGCTTTGTAGACAGATGAAGAAAGAT 720
   |||||
Db 232 ProCysAlaValaValaSerProGlnMetIleGlnMetLysCysAspSerAspAlaLysHis 251
   |||||

QY 721 GCGTGTCTTGGCTGTGTGTGACACTACGATCATATGAGGCGCTGATGTGTGATCTG 780
   |||||
Db 252 SerLeuLeuGlnAlaGlyGlyGlyIleHisAlaArgIlePheGlyProAspGlySerAspLeu 271
   |||||

```

```

QY 781 GTCGGCCCTTTGCCGAAATGAAGAGGATTCTCTACGCAACCTTGATCTGGAGTA 840
   |||||
Db 272 AlatrProLeuGlyGlnHisGlyGlnGlyLeuLysTyrAlaThrLeuAspProAlaIle 291
   |||||

QY 841 GCGATCTCTTGAATATGCGCGACACCTGCTGCTCATTTATCCGCTCCGACATTA 900
   |||||
Db 292 LeuIlePheAlaLysValAlaIleAspProAlaGlyHisTyrSerThrArgProAspValThr 311
   |||||

QY 901 CGCTTGCTAATATGATGACGACCTTAATTAATCCGATGTTGAATGAAGCT 951
   |||||
Db 312 ArgLeuMetPheAsnProAsnProAsnProCysValValGlnLeuProGly 328
   |||||

RESULT 6
Q88B32_9Z2ZZZ PRELIMINARY; PRT; 333 AA.
ID Q88B32_9Z2ZZZ PRELIMINARY; PRT; 333 AA.
AC Q88B32_9Z2ZZZ PRELIMINARY; PRT; 333 AA.
DT 05-JUL-2004 (TRMBLrel. 27, Created)
DT 05-JUL-2004 (TRMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TRMBLrel. 27, Last annotation update)
DE Nitrilase (EC 3.5.5.7).
GN Nitrilase=BD7615;
OS unclassified organism.
OC unclassified; environmental samples.
OX NCBI_TaxID=155900;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX PubMed=15066841; DOI=10.1128/AEM.70.4.2429-2436.2004;
RA Robertson D.E., Chaplin J.A., Desantis G., Podar M., Madden M.,
RA Chi E., Richardson T., Miln A., Miller M., Weiner D.P., Wong K.,
RA McQuaid J., Farwell B., Preston L.A., Tan X., Sneed M.A., Keller M.,
RA Machur B., Kretz P.L., Burk M.J., Short J.M., Sneed M.A., Keller M.,
RA "Exploring nitrilase sequence space for enantioselective catalysis.";
RA Appl. Environ. Microbiol. 70:2429-2436(2004).
DR EMBL; AY487492; AAR97439.1; -; Genomic DNA.
DR GO; GO:0016810; F:aliphatic nitrilase activity; IEA.
DR GO; GO:0016810; F:hydrolyase activity, acting on carbon-nitrog. . .; IEA.
DR GO; GO:0006807; P:nitrogen compound metabolism; IEA.
DR InterPro; IPR000132; N/CN hydrolase.
DR InterPro; IPR003010; Ntase/CNhydase.
DR Pfam; PF00795; CN hydrolase; 1.
DR PROSITE; PS50263; CN HYDROLASE; 1.
DR PROSITE; PS00920; NITRIL_CHT_1; 1.
DR PROSITE; PS00921; NITRIL_CHT_2; UNKNOWN_1.
KW Hydrolase.
SQ SEQUENCE 333 AA; 36342 MW; B0A9B825FD0D7403 CRC64;


Alignment Scores:



| Pred. No.:             | 4.31e-93                                  | Length:       | 333 |
|------------------------|-------------------------------------------|---------------|-----|
| Score:                 | 1139.00 <td>Matches:</td> <td>212 </td>   | Matches:      | 212 |
| Percent Similarity:    | 79.0% <td>Conservative:</td> <td>36 </td> | Conservative: | 36  |
| Best Local Similarity: | 67.5% <td>Mismatches:</td> <td>66 </td>   | Mismatches:   | 66  |
| Query Match:           | 62.4% <td>Indels:</td> <td>0 </td>        | Indels:       | 0   |
| DB:                    | 2                                         | Gaps:         | 0   |



US-09-751-299-3 (1-1014) x Q88B32_9Z2ZZZ (1-333)



```

QY 1 ATGAAAGAGCTATCAAGTCCCTGCGTGAAGCCCGCATCTACATGATTGGAG 60
   |||||
Db 2 MetlysaLProleuylValAlaCysValGlnAlaIleProMetPheLeuAsnLeuasp 21
   |||||

QY 61 GCGACGGTGGACAAACCTTGAATTGATGGAAGACGACGCTAATATGCTGCTG 120
   |||||
Db 22 GlyThrIleAspLysThrIleThrLeuIleSerGlnAlaIleGlnLysGlyAlaGlnLeu 41
   |||||

QY 121 ATGCGCTTCCGGAACCTTGATTCAGGCTACCCATGGTCTTCTGGCTGATCACA 180
   |||||
Db 52 IlealapeProGlnThrIleProGlyIleProTyrPheLeuThrPheLeuAspIlePro 61
   |||||

QY 181 GCATGGGCAATGCAATTTGTAGCCCAATACCATGAGAACTCATGAGTTGAGTGCCT 240
   |||||
Db 72 AlatrPasnMetProLeuValTyrGlnTyrHisGlnGlnSerLeuValLeuAspSerAla 81
   |||||

```


```

```
QY 241 CAAGCTAAGCGCATTTGAGTACGAGCCGAAGCGGTTGGGATCATGTGACCCCTGGGAGATG 300
Db 82 GlnAlaYsArGllleAlaAspAlaAlaGlnAlaGlnsnnleThrValValLeuGlyPhe 101
QY 301 AGTGAACGGGTCGGTGGCAGCCCTTACATCAGTACAGTGGTTCATAGGCGGATTAAGTGAC 360
Db 102 SerGlnuArgAspHlsGlySerLeuTyrlleSerGlnTrpLeuIleGlySerAspGlyGlu 121
QY 361 ACCATTGGGGCCCGGCAAAATTGAAACCTACTTTGTGTAACGTAAGCTTGTGGCGAA 420
Db 122 ThrIleGlylleaArgGlyLeuLysAlaThrhlsValGlnArgThrLeuPheGlyGlu 141
QY 421 GGGGATGGTTCATGCTAGCGCGTTTTCGAGACGTCGTGGAGAGCGTGGGCTTATAGC 480
Db 142 SerAspGlySerSerLeuThrThrTrpGlnuThrProLeuGlyAsnValGlyAlaLeuGly 161
QY 481 TGTGGGAGACCTTCAACCGCCTAACAAATACGTTTGATGCAAAATGAAAGATT 540
Db 162 CysTrpGlnHlsLeuGlnProLeuSerArgTyrlaMetTyrsGlnHlsGlyGluIle 181
QY 541 CATGTGGCGGCTGGCGGAGCTTATAGCCTTATCCTTAATGGCGGAAAGCCCTGGGCGCT 600
Db 182 HlsIleAlaAlaTrpProSerPheSerLeuTyrlThrSerAlaThrAlaAlaLeuGlyPro 201
QY 601 GATGTCAATGTAGCGGCGCTTCGAATCTATGCGGTGAGGCAATGCTTGCTACTACG 660
Db 202 AspAlaAsnThrAlaAlaSerArgLeuTyrlaAlaGlnGlyGlnCysPheValLeuAla 221
QY 661 TCGTGTCGCTCGCTTTCACATTCATGATGCTTGTGTGACAGATGAGAAAGCAT 720
Db 222 ProCysAlaValAlaSerAspGlnMetIleAspLeuLeuGlyProAspAspArgArg 241
QY 721 GCGTGTCTTCGGTGTGTGTGACACTCAGTATCATAGGCGCTGATGTGTGACTG 780
Db 242 AlaLeuLeuSerAlaGlyGlyHlsAlaArgIleTyrlArgProAspGlyArgGlnLeu 261
QY 781 GTCCGCGCTCTTGGCGGAAATGAAAGGGTATTCTTACGCAAACTTGATCTGGAGTA 840
Db 262 ValThrProLeuGlyGlnAsnGlyGlnGlyLeuLeuIleAlaGlnLeuAspSerAlaAla 281
QY 841 CGCATCTTGTCTTAATAGGCGGAGACCTGCTGCTTATTCGCCGTCGCACTTACT 900
Db 282 IleThrPheAlaYsLeuAlaAlaAspProValGlyHlsTyrlSerArgProAspValThr 301
QY 901 CGCTTGCTAATATGCGGCGCTTAATTAACCGGTAGTTGAA 942
Db 302 ArgLeuLeuPheAsnProSerAlaAsnTyrlThrValIleGly 315
RESULT 7
Q6RMJ0_92222 PRELIMINARY; PRT; 332 AA.
ID Q6RMJ0_92222
AC Q6RMJ0;
DT 05-JUL-2004 (TREMBLrel. 27, Created)
DT 05-JUL-2004 (TREMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TREMBLrel. 27, Last annotation update)
DE NitriLase (EC 3.5.5.7).
ORFNames=SDS168;
OS unclassified organism.
OC unclassified; environmental samples.
CX NCBI_TaxID=155900;
RN [1]
RP NUCLEOTIDE SEQUENCE.
RX PubMed=15066841; DOI=10.1128/AEM.70.4.2429-2436.2004;
RA Robertson D.E., Chaplin J.A., Desantis G., Podar M., Madden M.,
RA Chl E., Richardson T., Milan A., Miller M., Weiner D.P., Wong K.,
RA McQuaid J., Fairwell B., Preston L.A., Tan X., Sneed M.A., Keller M.,
RA Machur E., Kretz P.L., Burk M.J., Short J.M.;
RT "Exploring nitriLase sequence space for enantioselective catalysis.";
RL Appl. Environ. Microbiol. 70:2429-2436(2004).
DR EMBL; AY487512; AAK97459.1; -; Genomic_DNA.
DR GO; GO:0016872; F:aliphatic nitriLase activity; IEA.
DR GO; GO:0016810; F:hydrolase activity; acting on carbon-nitrog. .; IEA.
DR GO; GO:0006807; P:nitrogen compound metabolism; IEA.
```

```
DR InterPro; IPR000132; N/CN hydratase.
DR InterPro; IPR003010; N/Lse/CNhydase.
DR Pfam; PF00795; CN_hydrolase; 1.
DR PROSITE; PS50263; CN_HYDROLASE; 1.
DR PROSITE; PS00920; NITRIL_CHT_1; 1.
DR PROSITE; PS00921; NITRIL_CHT_2; UNKNOWN_1.
KW HydroLase.
SQ SEQUENCE 332 AA; 36214 MW; 2E7B8470F426184 CRC64;
Alignment Scores:
Pred. No.: 8e-93 Length: 332
Score: 1136.00 Matches: 211
Percent Similarity: 79.0% Conservative: 37
Best Local Similarity: 67.2% Mismatches: 66
Query Match: 62.2% Indels: 0
DB: 2 Gaps: 0
US-09-751-299-3 (1-1014) x Q6RMJ0_92222 (1-332)
QY 1 ATGAAGAAGCTATCAAGTTCGCTGCGTCAAGCCCGCATCTACATGATTTGGAG 60
Db 1 MetLysThrThrValThrValAlaCysValGlnAlaAlaProValPheMetAspLeuGlu 20
QY 61 GCGACGGTGCAAAACCAATTGACTGATGGAAGACGACGCTAATTAATGCTGCTG 120
Db 21 GlyThrValAspLysThrIleThrLeuIleSerGlnAlaGlnTyrlGlyAlaGlnLeu 40
QY 121 ATGCGCTTTCGGAACTGGATTCGACGGTACCCATGAGTTTCTTGCGCTTGAACA 180
Db 41 IleAlaPheProGlnuThrIleProGlyTyrlProIlePheLeuThrLeuAsnSerPro 60
QY 181 GCATGGGCAATGCAATTTGTAAGCCCAATACCATGAACTCATTTGAGTTGATGGCC 240
Db 61 AlaThrAsnMetProLeuValTyrlGlnTyrlHlsGlnAsnSerLeuValLeuAspSerThr 80
QY 241 CAAGCTAAGCGCATTTGAGTACGAGCCGTCGAGGCTTGGGAATCATGTGACCTTGGGATG 300
Db 81 GlnAlaYsArGllleAlaAspAlaAlaArgGlnAsnAsnleThrValAlaLeuGlyPhe 100
QY 301 AGTGAACGGGTCGGTGGCAGCCCTTACATCAGTACAGTGGTTCATAGGCGGATTAAGTGAC 360
Db 101 SerGlnuArgAspHlsGlySerLeuTyrlleAlaGlnTrpLeuIleGlySerAspGlyGlu 120
QY 361 ACCATTGGGGCCCGGCAAAATTGAAACCTACTTTGTGTAACGTAAGCTTGTGGCGAA 420
Db 121 ThrIleGlylleaArgGlyLeuLysAlaThrhlsValGlnArgThrLeuPheGlyGlu 140
QY 421 GGGGATGGTTCATGCTAGCGGTTTTCGAGACGTCGTGGAGAGCGTGGGCTTATAGC 480
Db 141 SerAspGlySerSerLeuThrThrTrpGlnuThrProLeuGlyAsnValGlyAlaLeuGly 160
QY 481 TGTGGGAGACCTTCAACCGCCTTAACAAATACGTTTGATGCAAAATGAAAGATT 540
Db 161 CysTrpGlnHlsLeuGlnProLeuSerArgTyrlaMetTyrsGlnHlsGlyGluIle 180
QY 541 CATGTGGCGGCTGGCGGAGCTTATAGCCTTATCCTTAATGCGGGAAGCCCTGGGCGCT 600
Db 181 HlsIleAlaAlaTrpProSerPheSerLeuTyrlThrSerAlaThrAlaAlaLeuGlyPro 200
QY 601 GATGTCAATGTAGCGGCGCTTCGAATCTATGCGGTGAGGCGCAATGCTTGCTACTACG 660
Db 201 AspAlaAsnThrAlaAlaSerArgLeuTyrlaAlaGlnGlyGlnCysPheValIleAla 220
QY 661 TCGTGTCGCTCGTTTCAACATTCATGATGCTTGTGTAACGTAAGCTTGTGGCGAA 720
Db 221 ProCysAlaValAlaSerAspGlnMetIleAspLeuLeuGlyProAspAspArgArg 240
QY 721 GCGTGTCTTCGGTGTGTGTGACACTCAGTATCATAGGCGCTGATGTGTGACTG 780
Db 241 AlaLeuLeuSerAlaGlyGlyHlsAlaArgIleTyrlArgProAspGlyArgGlnLeu 260
QY 781 GTCCGCGCTCTTGGCGGAAATGAAAGGGTATTCTTACGCAAACTTGATCTGGAGTA 840
```

Db 261 ValThrProLeuGlyGluGlnGluGlyLeuLeuIleAlaGluLeuLeuSerAlaAla 280  
QY 841 CGATCTCTGCTTAATAATGGGGGAGACCCCTGCTGTCATTATCCCGTCCCGCATTAAT 900  
Db 281 ILeThrPheAlaLeuIleGluGlnAlaAlaAspProValGlyHisIleTyrSerArgProAspValThr 300  
QY 901 CCGTCTGCTAATAGATCGACGCCCTTAATATTCACGGTGTGTTGA 942  
Db 301 ArgLeuLeuPheAsnProSerAlaAsnIleThrValIleIleLys 314  
  
RESULT 8  
Q6RWS5\_92ZZZ PRELIMINARY; PRT: 345 AA.  
ID Q6RWS5\_92ZZZ  
AC Q6RWS5;  
DT 05-JUL-2004 (TrEMBLrel. 27, Created)  
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)  
DE Nitrilase (EC 3.5.5.7).  
GN ORFNames=BD7611;  
OS uncultured organism.  
OC unclassified; environmental samples.  
OX NCBI\_TaxId=155900;  
RN (1)  
RP NUCLEOTIDE SEQUENCE.  
RX PubMed=15066841; DOI=10.1128/AEM.70.4.2429-2436.2004;  
RA Robertson D.E., Chaplin J.A., Desantis G., Podar M., Madden M.,  
RA Chi E., Richardson T., Milan A., Miller M., Weiner D.P., Wong K.,  
RA McQuaid J., Farwell B., Preston L.A., Tan X., Sneed M.A., Keller M.,  
RA Mathur E., Kretz P.L., Burk M.J., Short J.M.;  
RT "Exploring nitrilase sequence space for enantioselective catalysis.";  
RL Appl. Environ. Microbiol. 70:2429-2436(2004).  
DR EMBL: AY487427; FARS374.1; -; Genomic\_DNA.  
DR GO: 0018762; F:aliphatic nitrilase activity; IEA.  
DR GO: 0016810; F:hydrolase activity, acting on carbon-nitrog. . .; IEA.  
DR InterPro: IPR001807; P:nitrogen compound metabolism; IEA.  
DR InterPro: IPR003010; N:CN hydrolase.  
DR Pfam: PF00795; CN\_hydrolase; 1.  
DR PROSITE: PS50263; CN\_HYDROLASE; 1.  
DR PROSITE: PS00920; NITRIL\_CHT\_1; 1.  
DR PROSITE: PS00921; NITRIL\_CHT\_2; 1.  
KW Hydrolase.  
SQ SEQUENCE 345 AA; 37242 MW; 5F8A3CF530D21D CRC64;  
  
Alignment Scores:  
Pred. No.: 6.92e-88 Length: 345  
Score: 1081.00 Matches: 201  
Percent Similarity: 78.6% Conservative: 41  
Best Local Similarity: 65.3% Mismatches: 66  
Query Match: 59.2% Indels: 0  
Gaps: 0  
  
US-09-751-299-3 (1-1014) x Q6RWS5\_92ZZZ (1-345)  
QY 16 AAGTGGCGCTGGTGGAGAGCGCCCGCATCTACATGATTGAGGCGGATGACAA 75  
Db 9 ArgValAlaAlaValAlaGlnAlaIleProGluPheLeuAsnLeuGlnAlaThrValAspLys 28  
QY 76 ACCATTGATTGATGAGAAGAACAGACAGCTAATAATGCTGCTGATGCGCTTCGCGGA 135  
Db 29 ThrIleAlaLeuIleGluGlnAlaAlaArgGlyAlaSerLeuIleAlaPheProGlu 48  
QY 136 ACTTGATTCACGCTCAACCATGCTTTCTTGCTGACTACACGATGGGCAATGCA 195  
Db 49 ThrTrpIleProGlyTyrProTrpPheAlaTrpLeuGlyAlaProIleTrpGlyMetLys 68  
QY 156 TTGTAACGCATATACATGAAGCAATCATGAGTTGAGTGGCCCTCAAGCTAAGCGCAT 255  
Db 69 PheIleGlnAlaIleTyrHisAspAsnSerMetValIleAspGlyAlaGlnPheGluIle 88  
QY 256 TCAGATGACGACGAGCGTTGGGAATCATGATGTCACCTTCGGGATGAGGAACGGTCCGT 315  
Db 89 AlaGlnAlaIleAsnArgLysAsnIleThrValValLeuGlyPheSerGluLysAspAla 108

QY 316 GGCACCCCTTAATCATGATGATGCTTCATAGCGAATATGTCAGACCAATTGGGCGCCG 375  
Db 109 GlySerLeuTyrIleLeuGlnAlaIleLeuSerProGluGlyLysThrIleAlaThrArg 128  
QY 376 CGAAAGTTAAACCTACTTTTGTGAAAGCTACTTTGTTGGGGAAGGGATGGTTCATCG 435  
Db 129 ArgLysLeuLysProThrHisValGlnAlaArgAlaIlePheGlyGluLysAspLysSerAsp 148  
QY 436 CTAGCGGTTTTGAGAAGCTCTGTGTGGAAGGCTGGGTGGCTTAAGCTTTGGAGACCTT 495  
Db 149 LeuAlaValAlaHisAspThrLysLeuGlyArgValGlyAlaLeuGlyCysThrIleAlaIle 168  
QY 496 CAACCGCTTAACAAATACGCTTTGATAGCAAAATGAAGAATTCATGTGCGGCTGG 555  
Db 169 GlnProLeuSerLysTyrAlaMetTyrAlaGlnAsnGluGlnValHisIleAlaIleTrp 188  
QY 556 CCGAGCTTAGCCTTTATCTAATGCGGGAAGCGCTGGGCGCTGATGTCATGTAGCG 615  
Db 189 ProSerPheSerLeuTyrAlaAspAlaIleAlaIleAlaIleAlaIleAlaIleAlaIle 208  
QY 616 GCCTCTCGAATCTATGCGCTTGAAGGCAATGCTTCGTAAGCTGCTGCTGCT 675  
Db 209 AlaSerArgLeuTyrAlaValGluGlnGlyGlnCysPheValValAlaProCysAlaThrVal 228  
QY 676 TCACATCATCATATGATGATGCTTTGTCACAGTACGAAAGCAATGCTTCTTGCGCT 735  
Db 229 SerGlnLysMetIleAspMetLeuGlySerGluThrProGluGlnGlnAlaLeuLeuLysPro 248  
QY 736 GGTGTGTGACATCACTACATGATCAATAGGCGCTGATGCTGCTGCTGCTGCTGCTG 795  
Db 249 GlyLysGlyHisAspGlnIleTyrGlyProAspGlyArgSerLeuAlaAspProLeuPro 268  
QY 796 GAAATGAAGAGGTATTCTTACCCAAACCTTGATCTCTGAGTACGCATCTTGTCTTAA 855  
Db 269 ProAspAlaGluGlyLeuLeuLysTyrAlaAspIleAspLeuAlaAlaIleThrLeuAlaLys 288  
QY 856 ATGGCGGAGACCCCTGCTGCTGTCATTATTCGCTCCGACATTAATGCTGCTTAATGAT 915  
Db 289 AlaAlaAlaAspProAlaGlyHisIleTyrSerArgProAspValThrGlnLeuLeuAsp 308  
QY 916 CGCAGCCCTTAATATTCACGGTATTGTT 939  
Db 309 ArgAsnProLysProArgValVal 316  
  
RESULT 9  
Q6RWS3\_92ZZZ PRELIMINARY; PRT: 334 AA.  
ID Q6RWS3\_92ZZZ  
AC Q6RWS3;  
DT 05-JUL-2004 (TrEMBLrel. 27, Created)  
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)  
DE Nitrilase (EC 3.5.5.7).  
GN ORFNames=BD7051;  
OS uncultured organism.  
OC unclassified; environmental samples.  
OX NCBI\_TaxId=155900;  
RN (1)  
RP NUCLEOTIDE SEQUENCE.  
RX PubMed=15066841; DOI=10.1128/AEM.70.4.2429-2436.2004;  
RA Robertson D.E., Chaplin J.A., Desantis G., Podar M., Madden M.,  
RA Chi E., Richardson T., Milan A., Miller M., Weiner D.P., Wong K.,  
RA McQuaid J., Farwell B., Preston L.A., Tan X., Sneed M.A., Keller M.,  
RA Mathur E., Kretz P.L., Burk M.J., Short J.M.;  
RT "Exploring nitrilase sequence space for enantioselective catalysis.";  
RL Appl. Environ. Microbiol. 70:2429-2436(2004).  
DR EMBL: AY487559; FARS7506.1; -; Genomic\_DNA.  
DR GO: 0018762; F:aliphatic nitrilase activity; IEA.  
DR GO: 0016810; F:hydrolase activity, acting on carbon-nitrog. . .; IEA.  
DR InterPro: IPR001807; P:nitrogen compound metabolism; IEA.  
DR InterPro: IPR003010; N:CN hydrolase.  
DR Pfam: PF00795; CN\_hydrolase; 1.







```
Db      89  LysLeuArgAlaAlaAlaLysArgAsnLysVmeValValLeuLysLeuSerGluArg 108
Qy      310 GTCGGTGGACCCCTTACATCAAGTCAAGTTCATAGGGCAATAGTACACCATTTGG 369
Db      109 AaRGVLYLSeuLeuYrLleAlaGlnTrpLleGlyProAspLysGluThrIleAla 128
Qy      370 GCCCGGCGAAAGTTGAAACCTACTTTTGTGAACGTACTTTGTCGGCAAGGGATGGT 429
Db      129 LysArgArgLysLeuLysProThrHisAlaGluArgThValPheLysLysGluArgGly 148
Qy      430 TCATCGCTTACCGCTTTTGCAGACGTCTGTGGAAGGCTGGTGGCTTATGCTGTGGAG 489
Db      149 SerHisLeuAlaValHisGluLeuSerValGlyArgLysGluValAlaLeuLysCysTrpGlu 168
Qy      490 GACCTTCAACCGGTACAAATAACGCTTGTGTATGCAAAATGAGAGAGATTCATTGTGG 549
Db      169 HisLeuGlnProLeuSerLysTrpAlaMetTrpAlaGlnAsnGluGlnHisValAla 188
Qy      550 GCTTGGCCGAGCTTTTACCTTTTACCTTAATGCGGCGCAAAAGCCCTGGGCGCTGATCAAT 609
Db      189 AlaTrpProSerPheSerLeuYrAspProPheHisAlaLeuLysAlaGluValAsn 208
Qy      610 GTAAGCGCTCTGCAATCTATGCGCTTGAAGGCAATGCTTGTACTAGCGTGTGGG 669
Db      209 AaenAlaLaseLysLleTrpAlaValGluGlySerCysPheValIleAlaProCysAla 228
Qy      670 CTCGGTTTCAGATTCATGATGATATGCTTGTGTACAGATGAGAAAGATGCGTTCCT 729
Db      229 ThrValSerGlnAlaMetLleAspGluLeuCysAspThrProGluLysHisGlnPheLeu 248
Qy      730 CTGGCTGTGTGTGACACATCACTACATATGAGGCGCTGATGTGTGACTTGGTTCGGCCT 789
Db      249 HisAlaGlyLysGlyPheAlaValLleTrpGlyProAspGlyAlaLProLeuAlaAlaPro 268
Qy      790 CTTCGCCAAATGAAGAAGGATTTCTTACGCAAACTTGTATCTGAGATCAAGATCTT 849
Db      269 LeuProProAspLysGluLysLeuLysTrpAlaAspLysGluLysLleSerVal 288
Qy      850 GCTAAATGAGCGGACGACCCCTGCTGCTTATTCCTCCGCGCAATTAATCTGCTTGGTA 909
Db      289 AlaLysAlaAlaAlaAspProAlaGlyHisTrpAlaArgProAspValThrArgLeuLeu 308
Qy      910 ATAGATCGACGCTT 924
Db      309 PheAsnAsnArgPro 313

RESULT 11
Q6RMS2_92ZZZ
Q6RMS2_92ZZZ PRELIMINARY; PRT; 348 AA.
AC Q6RMS2;
DT 05-JUL-2004 (TREMBLrel. 27, Created)
DT 05-JUL-2004 (TREMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TREMBLrel. 27, Last annotation update)
DE Nitrilase (EC 3.5.5.7).
GN ORFNames=BD5338;
OS uncultured organism.
OC unclassified; environmental samples.
OX NCBI_TaxId=155900;
RN (1)
RP NUCLEOTIDE SEQUENCE.
RX PubMed=15066841; DOI=10.1128/AEM.70.4.2429-2436.2004;
RA Robertson D.E., Chaplin J.A., Desantis G., Podar M., Madsen M.,
RA Chi E., Richardson T., Milan A., Miller M., Weiner D.P., Wong K.,
RA McQuaid J., Ferwell B., Preston L.A., Tan X., Sneed M.A., Keller M.,
RA Mathur E., Kretz P.L., Burk M.J., Short J.M.;
RT "Exploiting nitrilase sequence space for enantioselective catalysis.";
RL EMBL; AY487430; AAR9377.1; 70:2429-2436(2004).
DR GO; GO:0018762; F:aphatic nitrilase activity; IEA.
DR GO; GO:0016810; F:hydrolase activity, acting on carbon-nitrog. . .; IEA.
DR InterPro; IPR000132; N/CN hydrolase.
DR InterPro; IPR003010; Ntlse/CNhydase.
```

```
DR InterPro; IPR002332; PII_GlnB_UMP_S.
DR Pfam; PF00795; CN_hydrolase_1.
DR PROSITE; PS00263; CN_HYDROLASE; 1.
DR PROSITE; PS00920; NITRIL_CHT_1; 1.
DR PROSITE; PS00921; NITRIL_CHT_2; UNKNOWN 1.
DR PROSITE; PS00496; PII_GlnB_UMP; UNKNOWN 1.
DR Hydrolase.
SQ SEQUENCE 348 AA; 38397 MW; 95F18795D55C01A5 CRC64;

Alignment Scores:
Pred. No.: 4,05e-84 Length: 348
Score: 1039.00 Matches: 191
Percent Similarity: 78.2% Conservative: 46
Best Local Similarity: 63.0% Mismatches: 66
Query Match: 56.9% Indels: 0
DB: 2 Gaps: 0

US-09-751-299-3 (1-1014) x Q6RMS2_92ZZZ (1-348)
Qy      16 AAGTGGCTGTGGTGGCAAGCCGCGCATCTACATGATTTGGAGCGACGTTGACAA 75
Db      8 ArgValAlaAlaValAlaGlnAlaAlaProValPheLeuAspLeuGluValAlaLleSerLys 27
Qy      76 ACCATTGAGTTGATGGAAGAAGACAGACAGTATAATGCTCGTGTGATGCGCTTTCGGAA 135
Db      28 GlyLleSerLeuLleGluGluAlaLaseHisGlnGlyAlaLysLeuLleAlaPheProGlu 47
Qy      136 ACTTGATTCAGAGCTACCCATGCTTCTTGGCTTGACTGACACGACATGGGCAATGCA 195
Db      48 ThrTrpLleProGlyTrpProTrpTrpLleTrpLeuAspSerProAlaTrpGlyMetArg 67
Qy      196 TTGTGACCCCAATCCATGAGAACATCATTTGAGAGTTGATGAGCCCTCAGCTTAAGGCAAT 255
Db      68 PheValGlnArgTrpPheAspAsnSerLeuMetLeuGlySerGluGlnAlaLysArgMet 87
Qy      256 TCAGATGAGCCCAAGCCGTTGGAATCATGCTGTCACCCCTGGGAGATGAGAAAGGCTCGGT 315
Db      88 AaenGlnAlaAlaAlaAsnAsnLysLleTrpValValMetGlyTrpSerGluArgSerLys 107
Qy      316 GGCACCTTTATCATCATGCTAGCTGCTTCAATAGCGATTAATGATGACACCATTTGGGCGG 375
Db      108 GlySerLeuYrMetGlyGlnSerLleAsnAspLysGlyGluThrLlePheThrArg 127
Qy      376 CGAAAGTTGAACCTACTTTTGTGAACGTACTTTGTCGGCGAAGGAGATGCTTCACT 435
Db      128 ArgLysLeuLysProThrHisValGluArgThrValPheGlyGluGlyAspGlySerHis 147
Qy      436 CTAGCGGCTTTTGCAGACGTCTGTGGAAGGCTGGGCTTATGCTGTGGAGACCTT 495
Db      148 LeuCysValMetAspThrGlnLleGlyArgValGlyAlaMetCysCysTrpGluHisLeu 167
Qy      496 CAACCGCTTACAAATAACGCTTGTATGACCAAAATGAGAGATTCATTGTGCGGCTTGG 555
Db      168 GlnProLeuSerLysTrpAlaMetLysSerGlnAspGluGlnLleHisLleLleSerTrp 187
Qy      556 CCGAGCTTTTACCTTTTATCTTAATGCGGCGCAAAAGCCCTGGGCGCTGATGATGACG 615
Db      188 ProSerPheSerLeuYrArgGlyAlaAlaLysTrpAlaLeuGlyProGluLeuAsnAsnAla 207
Qy      616 GCTCTCGAATCTATGCGTGTGGAAGGCAATGCTTGTGATGAGCGTGTGGCGCTTGT 675
Db      208 AlaSerGlnMetTrpAlaAlaGluGlyGlnCysPheValLeuAlaLProCysAlaThrVal 227
Qy      676 TCACATTCATGATGATGATGATGCTTGTGACAGATGACGAAAGATGCGTTCCTTGTGCT 735
Db      228 SerLysGluMetLleGluMetLeuLleAspAspProArgLysGluProLeuLeuLeuGlu 247
Qy      736 GGTGTGACACATCACTATATGAGGCTGATGCTGTGATGCTGTGCGGCTTGTGCC 795
Db      248 GlyGlyGlyPheThrMetLleTrpGlyProAspLysArgProLeuAlaLysProLeuPro 267
Qy      796 GAAATGAAGAGGATATCTCTACGCAAACTTATCTGAGTAACGATCTTGTCTTAA 855
```





```

QY 250 CGGATTTCAGATGACGCAAGCGGTGGATGATGTCACCTGGGAGATGATGAACGG 309
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
      89 LysLeuArgLeuAlaValLysLysAlaGlyLeuThrAlaValIleGlyLeuSerGluArg 108
QY 310 GTGCGTGGACCCCTTTCATCATGTCAGTGGTTCATAGCCGATATAGTGACACCATGGG 369
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
      109 GluGlyLysSerLeuTyrLeuAlaGlnTrpLeuIleGlyProAspGlyGluThrIleAla 128
QY 370 GCCCGGGAAGTGAAGCTACTCTTTGTTGACGTCATCTTTGCTGGCGGAAGGGATGGT 429
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
      129 LysArgAlaGlyLysLeuAlaGlyProThrHisAlaGluThrValTyrGlyGluGlyAspGly 148
QY 430 TCATCGCTAGCGGCTTTTCGAG--ACGTCCTGTTGAAAGCTGGGCTTATGCTGTGG 486
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
      149 SerAspLeuAlaValHisAspArgProGlyIleGlyArgLeuGlyAlaLeuGlyCysGlyTrp 168
QY 487 GAGGACCTTCAACCGCTTACAAATAAGCTTTGATGATGACAAATGGAAGATTCATTTGT 546
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
      169 GluHisLeuGlnProLeuSerLysTyrAlaMetTyrAlaGlnAsnGluGlnValHisVal 188
QY 547 GCGGCTGGCCGAGCTTTAGCCTTATCTTATCGGCGGAAGCCCTGGGCGCTGATGTC 606
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
      189 AlaAlaIleTrpSerPheSerLeuTyrAspProPheAlaProAlaLeuGlyTrpGluVal 208
QY 607 AATGTAGCGGCTCTCGAATCTATGCGCTTGAAGGCAATGCTTGTATAGCTGCTGT 666
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
      209 AsnAsnAlaIleSerArgValTyrAlaGluGlySerCysPheValLeuAlaProCys 228
QY 667 GCGCTGCTTCAACATCATGATGATGATGCTTGTATGATGATGATGATGATGATGATGATG 726
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
      229 AlaThrValSerLysAlaMetIleAspGluLeuGlyAspArgAspArgLysHisGlyLeu 248
QY 727 CTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 786
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
      249 LeuHisValGlyGlyHisAlaAlaIleTyrGlyProAspGlySerSerIleAlaGlu 268
QY 787 CCTCTGCGGGAATGGAAGAGGATTTCTTACGGAACCTGTAATCCCGAGTACGATC 846
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
      269 LysLeuProGluGlnGluGlyLeuTyrAlaAspIleAspLeuGlyAlaIleGly 288
QY 847 CTGCTAATAATGCGGCGAGCCCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 906
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
      289 IleAlaLysAsnAlaIleAspProIleGlyHisTyrSerArgProAspValThrArgLeu 308
QY 907 CTATATGATCCAGCCCT-----AAATTAACCGTATGTTGAATTT 945
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
      309 LeuLeuAsnLysLysProSerLysArgValGluHisPheSerLeuProValAspAsnVal 328
QY 946 GAAAGTATCTT 957
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
      329 GluProGluIle 332
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 15
GRRME4_92ZZZ PRELIMINARY; PRT: 338 AA.
AC Q6RME4_4 05-JUL-2004 (TrEMBLrel. 27, Created)
AC Q6RME4_4 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)
DE Nitrilase (EC 3.5.5.7).
GN ORFNames=BD5287;
OS unclassified organism.
OC unclassified; environmental samples.
OX NCBI_TaxID=155900;
RN NUCLEOTIDE SEQUENCE.
RX PubMed=1506841; DOI=10.1128/AEM.70.4.2429-2436.2004;
RA Robertson D.E., Chodari J.A., Desantis G., Podar M., Menden M.,
RA Chi E., Richardson T., Milan A., Miller M., Weiner D.P., Wong K.,
RA McQuaid J., Fairwell B., Preston L.A., Tan X., Sneed M.A., Keller M.,
RA Mathur E., Kretz P.T., Burk M.J., Short J.M.;
"Exploring nitrilase sequence space for enantioselective catalysis.";

```

```

RL Appl. Environ. Microbiol. 70:2429-2436(2004).
DR EMBL; AY487558; AAR97505.1; -; Genomic DNA.
DR GO; GO:0018762; F:aliphatic nitrilase activity; IEA.
DR GO; GO:0016810; F:hydrolase activity, acting on carbon-nitrog. . .; IEA.
DR GO; GO:0006807; P:nitrogen compound metabolism; IEA.
DR InterPro; IPR000132; N/CN hydrolase.
DR InterPro; IPR003010; Ntase/CNhydrlase.
DR Pfam; PF00795; CN_hydrolase; 1.
DR PROSITE; PS50263; CN_HYDROLASE; 1.
DR PROSITE; PS00920; NITRIL_CHT_1; 1.
DR PROSITE; PS00921; NITRIL_CHT_2; 1.
KW HYDROLASE.
SQ
SEQUENCE 338 AA; 36487 MW; 22F4109903344F59 CRC64;

Alignment Scores:
Pred. No.: 1.39e-83 Length: 338
Score: 1033.00 Matches: 194
Percent Similarity: 76.5% Conservative: 53
Best Local Similarity: 60.1% Mismatches: 66
Query Match: 56.6% Indels: 10
DB: Gaps: 3

US-09-751-299-3 (1-1014) x Q6RME4_92ZZZ (1-338)
QY 16 AAGGTGCGCTGCGGCAAGCGCCCGATCTACATGATTTGAGGCGGACGATGACAA 75
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
      9 ArgValAlaValAlaGlnAlaIleProAlaIleTrpLeuAspLeuGluAlaSerValSerLys 28
QY 76 ACCATTGATGTTAGGAAGACGACGATTAATCTCGTGTGATGCTTTCCGGA 135
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
      29 SerIleAlaLeuIleGluGluAlaAlaLysGlyAlaLysLeuIleAlaPheProGlu 48
QY 136 ACTTGATTCAGGCTACCCATGATGTTCTTGGCTTACTCACCAGCATGGCAATG--- 192
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
      49 AlaPheIleProGlyTyrProTrpTyrIleTrpLeuAspSerProAlaIleAlaIleGly 68
QY 193 ---CAATTTGACGCAATACATGAGACTCATTTGAGTGGATGGCCCTCAAGTAA 249
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
      69 ArgGlyPheValGlnArgTyrPheAspAsnSerLeuSerTyrAspSerProAlaIleGlu 88
QY 250 CGCATTTCAAGTACGACGCAAGCGGTGGGAATCATGTGACCCCTGGGAGATGAGAAC 309
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
      89 ArgLeuArgLeuAlaValLysLysAlaGlyMetThrAlaValLeuGlyLeuSerGluArg 108
QY 310 GTGCGTGGACCCCTTTCATCATGTCAGTGGTTCATAGCCGATATAGTGACACCATGGG 369
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
      109 AspGlyLysSerLeuTyrLeuAlaGlnTrpLeuIleGlyProAspGlyGluThrIleAla 128
QY 370 GCCCGGGAAGTGAAGCTACTCTTTGTTGACGTCATCTTTGCTGGCGGAAGGGATGGT 429
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
      129 LysArgAlaGlyLysLeuAlaGlyProThrHisAlaGluThrValTyrGlyGluGlyAspGly 148
QY 430 TCATCGCTAGCGGCTTTTCGAG--ACGTCCTGTTGAAAGCTGGGCTTATGCTGTGG 486
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
      149 SerAspLeuAlaValHisAspArgProGlyIleGlyArgLeuGlyAlaLeuGlyCysGlyTrp 168
QY 487 GAGGACCTTCAACCGCTTACAAATAAGCTTTGATGATGACAAATGGAAGATTCATTTGT 546
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
      169 GluHisLeuGlnProLeuSerLysTyrAlaMetTyrAlaGlnAsnGluGlnValHisVal 188
QY 547 GCGGCTGGCCGAGCTTTAGCCTTATCTTATCGGCGGAAGCCCTGGGCGCTGATGTC 606
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
      189 AlaAlaIleTrpSerPheSerLeuTyrAspProPheAlaProAlaLeuGlyTrpGluVal 208
QY 607 AATGTAGCGGCTCTCGAATCTATGCGCTTGAAGGCAATGCTTGTATAGCTGCTGCTGT 666
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
      209 AsnAsnAlaIleSerArgValTyrAlaGluGlySerCysPheValLeuAlaProCys 228
QY 667 GCGCTGCTTCAACATCATGATGATGATGCTTGTATGATGATGATGATGATGATGATGATG 726
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
      229 AlaThrValSerGlnAlaMetIleAspGluLeuGlyAspArgAspArgLysHisGlyLeu 248
QY 727 CTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 786

```

```

Db      249 LeuHisValGlyGlyHisAlaIlePheGlyProAspGlySerAlaIleAlaAsp 268
QY      787 CCTTGGCCGAAATGAGAGGCTATTCTTACGCAACCTTGATCTGGAGTACGCATC 846
Db      269 LysLeuProSerAspGlnGlnGlyLeuLeuPheAlaAspIleAspLeuGlyAlaIleGly 288
QY      847 CTTCCTAAATGCGCGCAGACCCCTGCTGTCATTAATCCCGTCCGACATTACTGCTTG 906
Db      289 IleAlaLysAsnAlaIleAspProIleGlyHisTyrSerArgProAspValThrArgLeu 308
QY      907 CTATATGATCGAGCCCTAAA-----TTACCGGTAGTTGAAATT 945
Db      309 LeuLeuAsnLysLysProSerLysArgValGluHisPheAlaLeuProLeuAspThrLeu 328
QY      946 GAAGGTGAT 954
Db      329 AlaGlyGlu 331

```

Search completed: April 27, 2006, 01:35:26  
 Job time : 256.429 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: April 27, 2006, 01:06:26 ; Search time 33.8987 Seconds  
(without alignments)  
2628.598 Million cell updates/sec

Title: US-09-751-299-3

Perfect score: 1826

Sequence: 1 atgaagaagcgtcatcaagtc.....cgcactcgagaatttga 1014

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 4886326

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODE=frame+npj.model -DEV=xlp  
-G=/abs/ABSSMBR.epool/US09751299/runat\_26042006\_090032\_18619/app.query.fasta\_1  
-DB=A.Geneseq -OFFMT=faстан -SUFFIX=rag -MINMATCH=0.1 -LOOPEL=0 -LOOPEXT=0  
-UNITS=bits -START=1 -END=-1 -MATRIX=blonum62 -TRANS=human40.cdi -LIST=45  
-LOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15 -MODE=LOCAL  
-OUTFMT=pco -NORM=ext -HEA\_SIZE=500 -MINLEN=0 -MAXLEN=200000000 -HOST=abs802p  
-USER=US09751299\_@CGN\_1\_1\_476\_@runat\_26042006\_090032\_18619 -NCPU=6 -ICPU=3  
-NO MMAP -NEG SCORES=0 -WAIT -DSBLOCK=100 -LONGLOG -DEV TIMEOUT=120  
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-XGAPOP=10 -XGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

A\_Geneseq\_21: \*  
1: geneseqp1980s: \*  
2: geneseqp1990s: \*  
3: geneseqp2000s: \*  
4: geneseqp2001s: \*  
5: geneseqp2002s: \*  
6: geneseqp2003as: \*  
7: geneseqp2003bs: \*  
8: geneseqp2004s: \*  
9: geneseqp2005s: \*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Query Length	DB ID	Description
1	1765	96.7	337	4	AAE05489 Nitriase
2	1761	96.4	337	7	Adc24119 Protein s
3	1761	96.4	337	8	Adh36220 Chemical
4	1761	96.4	337	8	Adg93919 Nitriase
5	1761	96.4	337	8	Adi62518 Nitriase
6	1761	96.4	337	8	Adi64639 Nitriase
7	1339	73.3	337	7	Adc23791 Protein s
8	1339	73.3	337	7	Adc23781 Protein s
9	1339	73.3	337	8	Adh35882 Chemical

10	1339	73.3	337	8	Adh35892 Chemical
11	1339	73.3	337	8	Adg93583 Nitriase
12	1339	73.3	337	8	Adg93593 Nitriase
13	1339	73.3	337	8	Adi62180 Nitriase
14	1339	73.3	337	8	Adi62190 Nitriase
15	1339	73.3	337	8	Adi64311 Nitriase
16	1339	73.3	337	8	Adi64301 Nitriase
17	1339	62.4	333	7	Adc24025 Protein s
18	1339	62.4	333	8	Adh36126 Chemical
19	1139	62.4	333	8	Adg93827 Nitriase
20	1139	62.4	333	8	Adi62424 Nitriase
21	1139	62.4	333	8	Adi64545 Nitriase
22	1136	62.2	332	7	Adc23873 Protein s
23	1136	62.2	332	8	Adh35974 Chemical
24	1136	62.2	332	8	Adg93675 Nitriase
25	1136	62.2	332	8	Adi62272 Nitriase
26	1136	62.2	332	8	Adi64393 Nitriase
27	1091	59.7	355	7	Adc24039 Protein s
28	1091	59.7	355	8	Adh36140 Chemical
29	1091	59.7	355	8	Adg93941 Nitriase
30	1091	59.7	355	8	Adi62438 Nitriase
31	1091	59.7	355	8	Adi64559 Nitriase
32	1081	59.2	345	7	Adc24067 Protein s
33	1081	59.2	345	8	Adh36168 Chemical
34	1081	59.2	345	8	Adg93968 Nitriase
35	1081	59.2	345	8	Adi62466 Nitriase
36	1081	59.2	345	8	Adi64587 Nitriase
37	1051	57.6	334	7	Adc23739 Protein s
38	1051	57.6	334	8	Adh35840 Chemical
39	1051	57.6	334	8	Adg93541 Nitriase
40	1051	57.6	334	8	Adi62138 Nitriase
41	1051	57.6	334	8	Adi64259 Nitriase
42	1042	57.1	353	7	Adc23945 Protein s
43	1042	57.1	353	8	Adh36046 Chemical
44	1042	57.1	353	8	Adg93747 Nitriase
45	1042	57.1	353	8	Adi62344 Nitriase

#### ALIGNMENTS

RESULT 1  
AAE05489  
ID AAE05489 standard; protein; 337 AA.  
XX  
AC AAE05489;  
XX  
DT 25-SEP-2001 (first entry)  
XX  
DE Nitriase protein #2.  
XX  
KW Nitriase; enantiomer; alpha-substituted carboxylic acid;  
KW stereoselective; hydrolysis; amino nitrite; cyanohydrin.  
XX  
OS Unidentified.  
XX  
PN WO200148175-A2.  
XX  
PD 05-JUL-2001.  
XX  
PF 29-DEC-2000; 2000WO-US035555.  
XX  
PR 29-DEC-1999; 99US-0173609P.  
PR 07-DEC-2000; 2000US-0254414P.  
XX  
PA (DIYE-) DIVERSA CORP.  
PA (MADD-) MADDEN D.  
XX  
PI Madden M, Weiner DP, Chaplin JA;  
XX WPI; 2001-465211/50.  
DR N-PSDB; AAD11167.  
XX  
PT Producing alpha-substituted carboxylic acid enantiomers by contacting

PT aldehyde or ketone with cyanide, ammonia compound or its salt or amine  
 PT and hydrolyzing the resulting amino nitrile or cyanohydrin with  
 PT nitrilase.

PS Claim 30; Page 86-87; 87pp; English.

XX The present invention relates to methods for producing an  
 CC enantiomerically pure alpha-substituted carboxylic acid. The method  
 CC involves connecting an aldehyde or ketone with a cyanide containing  
 CC compound and an ammonia-containing compound or an ammonium salt or amine,  
 CC and stereoselectively hydrolyzing the resulting amino nitrile or  
 CC cyanohydrin intermediate with a nitrilase or a polypeptide having  
 CC nitrilase activity. The present sequence is a nitrilase protein #2

XX Sequence 337 AA;

Alignment Scores:

Pred. No.: 6,87e-185 Length: 337  
 Score: 1765.00 Matches: 337  
 Percent Similarity: 100.0% Conservative: 0  
 Best Local Similarity: 100.0% Mismatches: 0  
 Query Match: 96.7% Indels: 0  
 DB: 4 Gaps: 0

US-09-751-299-3 (1-1014) x AAE05489 (1-337)

QY 1 ATGAAAGAGCTATCAGAGTGCCTGCGCAAGCGGCCGATCTACATGATTTGGAG 60  
 DB 1 MetysgsluhalieleyValAlaCysValGlnAlaAlaProIleTyrMechAspLeu 20  
 QY 61 GCGAGCTGGACAAACCATTTGATTTGATGGAAGACGACGTAATTAATGCTGCTG 120  
 DB 21 AlaThrValAspLysThrIleGluMetGluGlnAlaAlaArgAsnAlaArgLeu 40  
 QY 121 ATCGCTTCCGGAACCTTGATTCAGGCTAACCGATGTTCTTGGCTGACCA 180  
 DB 41 IleAlaProGluThrIlePheProGlyTyrProIlePheLeuTyrLeuAspSerPro 60  
 QY 181 GCATGGGCAATGCAATTTGTACGCCCATACATGAGAACTCATTTGGATGGCCCT 240  
 DB 61 AlaTrpAlaMetGlnPheValArgIleThrIleGlnAsnSerLeuGluLeuAspLysPro 80  
 QY 241 CAAAGTAAGCGCATTTTCAGATGACGCAAGCGGTTGGAAATCATGTCACTGGGGATG 300  
 DB 81 GlnAlaLysArgIleSerAspAlaAlaLysArgLeuGlyIleMetValThrLeuGlyMet 100  
 QY 301 AGTGAACGGGTGCGTGCACCTTATCATCATGATCATGATGATGATGATGATGATG 360  
 DB 101 SerGluArgValGlyGlyThrLeuTyrIleSerGlnTyrPheIleGlyAspAsnGlyAsp 120  
 QY 361 ACCATTGGGGCGCGGCAAGTTGAACCTTATTTGTTGAACGTAATTTGTTGGCGAA 420  
 DB 121 ThrIleGlyAlaArgArgLysLeuLysProThrPheValGluArgThrLeuPheGlyGlu 140  
 QY 421 GGGGATGGTTCATGCTGCGGTTTTCAGACGTCGTTGGAAAGCTGGGTGCTTATG 480  
 DB 141 GlyAspGlySerSerLeuAlaValPheGluThrSerValGlyArgLeuGlyGlyLeuGly 160  
 QY 481 TGTTCGAGCACTTCAACCGCTTAACTGCTTGTATGCAAAATGAAGAGATT 540  
 DB 161 CysTrpGluHisLeuGlnProLeuThrIleTyrAlaLeuTyrAlaGlnAsnGluIle 180  
 QY 541 CATGTGGGGCTTGGCGGCTTATGCTTATCTTATGCGGCAAGCCCTGGGGCT 600  
 DB 181 HisCysAlaAlaAlaTrpProSerPheSerLeuTyrProAsnAlaAlaLysAlaLeuGlyPro 200  
 QY 601 GATGTCATGTAGCGGCTTCGATCTATGCGGTGAAGGCAATGCTGCTATACG 660  
 DB 201 AspAlaAsnValAlaAlaSerArgIleTyrAlaValGluGlyGlnCysPheValLeuAla 220  
 QY 661 TCGTGTGCGCTGCTTTCAACATTCATGATGATGCTTTGTACGATGACGAAAGCAT 720  
 DB 221 SerCysAlaLeuValSerGlnSerMetIleAspMetLeuCysThrAspAspGlyHis 240

QY 721 GCGTTGCTTGGCTGGTGGTGGACACTACATCATAGGCGCTGATGTTGACTTG 780  
 DB 241 AlaLeuLeuLeuAlaGlyGlyGlyHisSerArgIleIleGlyProAspGlyLysAspLeu 260  
 QY 781 GTGCGGCTTTCGCGAAATGAAGAGGATTTCTCTACGCAACTGATCTGAGTA 840  
 DB 261 ValAlaProLeuAlaGluAsnGluGlyIleLeuTyrAlaAsnLeuAspProGlyVal 280  
 QY 841 CGCATCTTGTCTAAATGCGCGGACAGCCCTGCTGCTATTTCCGTCGCCAATTA 900  
 DB 281 ArgIleLeuAlaLysMetAlaAlaAspProAlaGlyHisTyrSerAspProAspIleThr 300  
 QY 901 CGCTTGCTAATGATCCGAGCCCTAAATTCGCGTACGTGAAATGGAAGTATCTT 960  
 DB 301 ArgLeuLeuIleAspArgSerProLysLeuProValValGluIleGluGlyAspLeuArg 320

RESULT 2

ADC24119  
 ID ADC24119 standard; protein; 337 AA.

AC ADC24119;

DT 18-DEC-2003 (first entry)

DE Protein sequence (SeqID 386) exhibiting nitrilase activity.

KW enzyme; nitrilase; nitrile; cyanohydrin; ammonia; biocatalyst;

XX enantiomer; chiral medicine.

XX Unidentified.

XX WO200300840-A2.

XX 03-JAN-2003.

XX 15-MAY-2002; 2002WO-US015983.

XX 21-JUN-2001; 2001US-0300189P.

XX 30-JUL-2001; 2001US-0309006P.

XX 22-JAN-2002; 2002US-0351336P.

XX (DIVE-) DIVERSA CORP.

XX (MADD/) MADDEN D.

XX Madden M, Desantis G, Chaplin JA, Weiner DP, Milan A, Chi E;

XX Short JM, Burk M;

XX WPI; 2003-201417/19.

XX N-PSDB; ADC24118.

XX Novel nitrilase polypeptide, useful for making (R)- or (S)-ethyl-4-cyano-

PT 3-hydroxybutyric acid or (R)- or (S)-mandelic acid or (S)- or (R)-phenyl

PT lactic acid derivative and for producing pharmaceutical composition, and

PT food additive.

XX Claim 40; SEQ ID NO 386; 560pp; English.

XX This invention relates to nitrilases and the nucleic acids that encode  
 CC these enzymes thereof. Specifically, it refers to polypeptides that  
 CC exhibit nitrilase activity, i.e. the ability to directly hydrolyse  
 CC nitriles or cyanohydrins into their corresponding carboxylic acids and  
 CC ammonia. Nitrilases have commercial utility as biocatalysts for use in  
 CC the synthesis of enantiomerically pure aromatic and aliphatic amino  
 CC acids, as well as hydroxy acids, which are important for the development  
 CC of chiral medicines. Furthermore, the present invention describes  
 CC nitrilases, isolated from mesophilic microorganisms, that have improved  
 CC activity and stability at increased pH and temperature. They are also  
 CC inexpensive, efficient catalysts, have broad substrate specificity and



CC are capable of chiral differentiation. This polypeptide is a protein  
CC sequence that exhibits nitrilase activity of the invention.

XX Sequence 337 AA;

SO

Alignment Scores:  
Pred. No.: 1.9e-184 Length: 337  
Score: 1761.00 Matches: 336  
Percent Similarity: 100.0% Conservative: 1  
Best Local Similarity: 99.7% Mismatches: 0  
Query Match: 96.4% Indels: 0  
DB: 7 Gaps: 0

US-09-751-299-3 (1-1014) x ADC24119 (1-337)

```
QY 1 ATGAAGAAGCTATTAAGTGCCTGCGTGCAGCCGCCGATCTACATGATTTGGAG 60
Db 1 Metlysglualallelyvalalacyvalglnalalaproilletymetaspheulys 20
QY 61 GCGACGGTGGCAAAACATTGAGTTGATGGAAGAAGACGCTAATTAAGTCTGCTG 120
Db 21 Alathrvalasplsthrilleelueumecuglualaalargasnasnalargleu 40
QY 121 ATCCGCTTCCGAAACTTGATTCAGGCTACCCATGAGTTCTTGCTGACTCACCA 180
Db 41 Ilealpheprogluthtripileprogltyrprotpheleuttripleuspserrpro 60
QY 181 GCATGGGCAATGCAATTTGTATGCCAATACATGAAATCTATTGAGTTGGATGGCCT 240
Db 61 Alatrpalametglnphevalarglntyrhlsigluanserserleugluleusplpyro 80
QY 241 CAAGCTAAGCCGATTTGATGACGCAAGCCGTTGGGAATCATGCTACCTGGGGATG 300
Db 81 Glnalalyasrglleseraspalaalalyasargleuglyllemevalthreulymet 100
QY 301 AGTGAACGGGTGCGTGACACCTTTACATCAAGTCAAGTTCATACAGGCATATGTTAC 360
Db 101 Sergluargvalglglythreulryrilleserlintphneilleglyaspnasndlyasp 120
QY 361 ACCATTGGGGCCCGCGAAGTTGAACCTTACTTTGTTGAACTGTTGTCGGCGAA 420
Db 121 Thrilleglyalalargarglyleulysprothrphelvalgluargthreupheglylu 140
QY 421 GGGGATGTTTATCGCTACGCGTTTTCAGACGCTGTTGGAAGGCTGGGTGGCTTAAAC 480
Db 141 Glyaspglyserserleualavalphegluthtservalgllyargleuglyglyleucys 160
QY 481 TGTGGGAGCACTTCACACCGCTAACCAAAATACGCTTGTATGCAAAATGAAGAGATT 540
Db 161 Cystirpgluhsleuglnproleuthrlystlyralaleulryralaglnasnglulile 180
QY 541 CATTGTGGCGGCTGGCCGAGCTTTAGCCTTTATCCTTAATGCGCGCAAAAGCCCTGGGCGCT 600
Db 181 HisCysalalalatrproserpheaserleulryproasnasnalalalyalaleuglypro 200
QY 601 GATGTCAATGTAAGCGGCTCTTCGAATCTATGCGCTTGAAGGGCAATGCTTGTACTACCG 660
Db 201 Aspyalasnvalalalaserargllyerlavalgluglincysphelvalleuala 220
QY 661 TCGTGGCGCTCGTTTCAATCCATCCATGATGATGCTTTGTAGACGAGAAAGAGAT 720
Db 221 SerCysalaleuvalserglinsermelleasphelneuCystrnaspsglulyshts 240
QY 721 GCGTGTCTTGTGGTGTGTGTGACACTCACGTATCATAGGCGCTGATGATGTGACTTG 780
Db 241 Alaleuuleuualaglyglylyhserserhrglilleglyproasnglyglyaspheul 260
QY 781 GTCCGGCTCTTGGCGGAAATGAAGAAGGTATTTCTACGCAAACTTGATCTTGAAGTA 840
Db 261 Valalaproleuualagluasnuglulglylleulryralaasnleuaspsproglyval 280
QY 841 CGCATCTTGTGTAATAATGGCGGACAGCCGTGCTCATTAATTCGCCGCGACATTACT 900
```

```
Db 281 ArgilleuualalyMetelalalaspbroalaglyhstlyrserrargproaspllethr 300
QY 901 CGCTTGCAATTAAGATTCGAGCCCTAAATTACCGGATTTGAATGAAGGTGATCTTGT 960
Db 301 Argleuuleuleasparserprolyseulprovalvalglulileglyaspheularg 320
QY 961 CCTTACGCTTGGGTGAAGCGTCTGAGACGGGTCGCGCAACTCGAAGAAATT 1011
Db 321 Protyralaleuglylyalaserchulthrelyalaglnleuglulile 337

RESULT 3
ADH36220
ID ADH36220 standard; protein; 337 AA.
XX
AC ADH36220;
XX
DT 11-MAR-2004 (first entry)
XX
DE Chemical process monitoring-related nitrilase protein sequence SeqID386.
XX
KM chemical process monitoring; biochemical process monitoring; cyanide;
XX high throughput system; enzyme.
XX
OS Unidentified.
XX
PN WO2003098187-A2.
XX
PD 27-NOV-2003.
XX
PF 15-MAY-2003; 2003WO-US015639.
XX
PR 15-MAY-2002; 2002US-0380737P.
XX
PA (DIVE-) DIVERSA CORP.
XX
PI Weiner D, Chaplin JA, Chi E, Milan A, Desautels G, Burk MJ;
PI Mcquaid J, Stege J;
XX
DR WPI; 2004-142708/14.
XX
DR N-PSDB; ADH36219.
XX
PT Monitoring a chemical or biochemical process comprises providing a
PT reactant comprising a cyanide or a material that can be converted to
PT cyanide or a reactant that generates a cyanide or a material that can be
PT converted to cyanide.
XX
PS Claim 73; SEQ ID NO 386; 277bp; English.
XX
CC This invention relates to a novel method of monitoring chemical or
CC biochemical processes. The method involves providing a reactant
CC comprising cyanide (or a material that can be converted to a cyanide)
CC that generates as a reaction product cyanide or a material that can be
CC converted to cyanide and measuring the concentration of produced cyanide.
CC The method is useful for monitoring a chemical or biochemical process.
CC The method is effective for high throughput systems and is sufficiently
CC sensitive to detect a small amount of product. The present sequence is
CC that of a nitrilase protein which may be used in the method of the
CC invention.
XX
SQ Sequence 337 AA;
XX
XX
Alignment Scores:
Pred. No.: 1.9e-184 Length: 337
Score: 1761.00 Matches: 336
Percent Similarity: 100.0% Conservative: 1
Best Local Similarity: 99.7% Mismatches: 0
Query Match: 96.4% Indels: 0
DB: 8 Gaps: 0

US-09-751-299-3 (1-1014) x ADH36220 (1-337)
QY 1 ATGAAGAAGCTATTAAGTGCCTGCGTGCAGCCGCCGATCTACATGATTTGGAG 60
|||
```

Db 1 Metlysgluatlailelyvalalacyvalalinalaaproilerymetaspleuys 20  
QY 61 GCGACGGTGGCAAAACCTTGAAGTGAAGAAGACGATTAATGCTCGCTG 120  
Db 21 Alathrvalasplystrillegluleuwetluglualalaagaasnaalaaargleu 40  
QY 121 ATGCGCTTTCGGAACCTTGATTCAGAGCTACCCATGCTTCTTGGCTTGACTCACA 180  
Db 41 IlealapherprogluthrtripleProdiLytrProtrPhleuEntripleuAspSerPro 60  
QY 181 GCATGGCAATGCAATTTGTACGCCAATACCATGAACTCATTTGAGTGGAGCCCT 240  
Db 61 AlatrpalametcinphevalarglIntyrhlsiglunsserleuGluLeuAspGlyPro 80  
QY 241 CAACCTTAAGCGCATTTTCAATGACGAGCGGTTGGGAATCATAGTGCACCCCTGGGAGT 300  
Db 81 GlnalalyarglileseraspalalalyargleuglyllemevalthleuGlymet 100  
QY 301 AGTGAACGGGTGGGTGGACCCCTTACATCAGTCACTGATGTTCATAGGCGATTAATGTGAC 360  
Db 101 SerGluarGValgllyGlyThrleuTyrilleserGlntrPheileGlyAspAsnGlyasp 120  
QY 361 ACCATTGGGGCCCGGCAAAAGTTGAACCTTCTTGTGAACGTAAGTCTTGTGGCGAA 420  
Db 121 ThrilleGlyalaaargllyleuLysProthrPheValGluArgThrleuPheGlyGlu 140  
QY 421 GGGGATGGTTCATGCGTACGCTTTTTCAGACGTCGTGTGGAAGGCTGGGTGCTTATGC 480  
Db 141 GlyaspGlyserSerleuAlaValaphelGlnthrservalGlyhargleuGlyGlyLeuGly 160  
QY 481 TGTGGGAGACCTTCAACCGCTTAACAAATACGCTTGTATGCAAAATGAAGAGATT 540  
Db 161 CysTrpGlnHlsleuGlnProleuThrlystrlyalaleuTyrAlaGlnAsnGluGlu 180  
QY 541 CATGTGGGGTGGTGGCGAGCTTACGCTTATACCTTAATGCGGGAAGCCCTGGGGCT 600  
Db 181 HisCysalaaalatrProserPheSerleuTyrProAsnaalalalyalaleuGlyPro 200  
QY 601 GATGTCAATGTAGCGGCTCTCGAATCTATGCGCTTGAAGGCAATGCTGTACTACG 660  
Db 201 AspValaenValalalaserArglileTyrAlaValGluGlyIncYcPheValleuAla 220  
QY 661 TCGTGTGGCTCGCTTTCACATCCATGATCGATATGCTTGTGACATGACGAAAGCAT 720  
Db 221 SerCysalaleuValaserGlnserMetileaspMetleuCystrInAspAspGluLysHis 240  
QY 721 GCGTGTCTTGGCGGTGGTGGACACTCAGCTATCATAGGCGCTGATGGTGGTACTTG 780  
Db 241 AlaleuLeuLeuAlaGlyGlyGlyHlsSerArglleileGlyProAspGlyGlyaspLeu 260  
QY 781 GTGCGGCTCTTTCGGAATGAAGAGGATTTCTTACGCAACCTTGATCTTGAGTA 840  
Db 261 ValalAProleuAlaGluAsnGluGlyleuTyrAlaAenLeuAspProGlyVal 280  
QY 841 CGCATCTTGTGTAATGGCGGAGACCGCTGCTGCTTATCCCGTCCGAGCATTAAT 900  
Db 281 ArgileuAlaLysMetAlaAlaAspProAlaGlyHlsTyrSerArgProAspLleThr 300  
QY 901 CGCTTGAATAGATCGAGCGCTTAATTAACCGAGTGTGAATTAAGAGTGTGATCTTGT 960  
Db 301 ArgileuLeuIleAspArgSerProLysleuProValGluGluileGluGlyaspLeuArg 320  
QY 961 CTTTACGCTTTGGGTAAAGCGTCTGAGACGGGTGGCAACTGAGAAATTT 1011  
Db 321 ProTyrAlaleuGlyLysAlaserGlnthrglyalaglnleuGluGluile 337

RESULT 4  
ADG93919  
ID ADG93919 standard; protein, 337 AA.  
AC ADG93919;  
XX  
XX  
DT 11-MAR-2004 (first entry)

XX DE Nitriase enzyme amino acid sequence SeqID386.  
XX KM nitriase; nitrite; carboxylic acid; chemical process; pH; temperature;  
XX KM enantioselective transformation; enzyme.  
XX OS Unidentified.  
XX PN MO2003097810-A2.  
XX PD 27-NOV-2003.  
XX PE 15-MAY-2003; 2003MO-US015712.  
XX PR 15-MAY-2002; 2002US-00146772.  
XX PR 09-SEP-2002; 2002US-00241742.  
XX PA (DIVE-) DIVERSA CORP.  
XX P1 Desantis G, Short JM, Burk M, Wong K, Farwell R, Chatman K;  
XX DR MPI: 2004-090637/09.  
XX DR N-PSDB: ADG93918.  
XX PT New isolated or recombinant nucleic acid encoding a polypeptide having  
XX PT nitriase activity, useful for screening enantioselective transformation.  
XX PS Claim 44; SEQ ID NO 386; 295pp; English.  
CC This invention is related to a novel isolated or recombinant nucleic acid  
CC encoding a protein having nitriase activity. Nitriase's are capable of  
CC converting nitrite's directly to carboxylic acids and have great  
CC potential for use in industrial chemical processes. The isolated  
CC nitriase proteins of the invention have increased activity and stability  
CC at increased pH and temperature when compared to those conventionally  
CC used. In addition, the nucleic acid of the invention is useful for  
CC screening enantioselective transformation. The present sequence is that  
CC of a nitriase enzyme of the invention.  
XX SQ Sequence 337 AA;  
XX  
XX Alignment Scores:  
XX Pred. No.: 1.9e-184 Length: 337  
XX Score: 1761.00 Matches: 336  
XX Percent Similarity: 100.0% Conservative: 1  
XX Best Local Similarity: 99.7% Mismatches: 0  
XX Query Match: 96.4% Indels: 0  
XX DB: 8 Gaps: 0  
XX  
XX US-09-751-299-3 (1-1014) x ADG93919 (1-337)  
QY 1 ATGAAGAAGCTATCAAGTGCCTGCGTGACCGCCGATCTACATGATTTGGAG 60  
Db 1 Metlysgluatlailelyvalalacyvalalinalaaproilerymetaspleuys 20  
QY 61 GCGACGGTGGCAAAACCTTGAAGTGAAGAAGACGATTAATGCTCGCTG 120  
Db 21 Alathrvalasplystrillegluleuwetluglualalaagaasnaalaaargleu 40  
QY 121 ATGCGCTTTCGGAACCTTGATTCAGAGCTACCCATGCTTCTTGGCTTGACTCACA 180  
Db 41 IlealapherprogluthrtripleProdiLytrProtrPhleuEntripleuAspSerPro 60  
QY 181 GCATGGCAATGCAATTTGTACGCCAATACCATGAACTCATTTGAGTGGAGCCCT 240  
Db 61 AlatrpalametcinphevalarglIntyrhlsiglunsserleuGluLeuAspGlyPro 80  
QY 241 CAACCTTAAGCGCATTTTCAATGACGAGCGGTTGGGAATCATAGTGCACCCCTGGGAGT 300  
Db 81 GlnalalyarglileseraspalalalyargleuglyllemevalthleuGlymet 100  
QY 301 AGTGAACGGGTGGGTGGACCCCTTACATCAGTCACTGATGTTCATAGGCGATTAATGTGAC 360

```
Db      101  SerGIuArgValGIyGIYThrLeuYrIleSerGIInTrpHeIleGIyAePaenGIyAsp 120
QY      361  ACCATTGGGGCCGGGGAAGTTGAACCTATTGTTGAACGTAATTGTTGGGGA 420
Db      121  ThrIleGIyAlaIArgIArgIyLeuYsPProThrPheValGIuArgIThrLeuPheGIyGIu 140
QY      421  GGGGATGGTTCATCGGTAGCGGTTTTCAGACGCTCTGTTGAAGGCTGGGTGCTTAATGC 480
Db      141  GIyAspGIYSerSerIleuAlaValPheGIuThrSerValGIYArgIleuGIyGIYLeuCYs 160
QY      481  TGTGGGAGACCTTCAACCGCTTAACCAAAATACGCTTTGTATGCACAAAATGAAGATT 540
Db      161  CYsTrpGIuHISleuGIInProlLeuThrIYrYrAlaLeuYrAlaGIInaenGIuGIuIle 180
QY      541  CATTTGGCGCTTGGCGGAGCTTACCTTATCCCTTAATGGCGGGAAGCCCTGGGGCCT 600
Db      181  HIsCYsAlaIaIaIArgIProSerPheSerLeuYrProAsnAlaIaIySAlaLeuGIYPro 200
QY      601  GATGTCAATGTAGCGGCTCTCGAATCTATGCGGTTGAAGGGCAATGCTTCGTACTAGCG 660
Db      201  AspValAsnValAlaIaIaSerArgIleYrYrAlaValaGIuGIYInCYsPheValaLeuAla 220
QY      661  TCGTGTGGCGCTCGTTTCACAATCATGATCATGATATGCTTTGTATCAGATGACGAAAACAT 720
Db      221  SerCYsAlaLeuValSerGIInSerMetIleAspMetLeuCYsThrAspArgIuYsHIs 240
QY      721  GGGTTCCTTCTGGCTGGTGGTGGACCTACGATATATAGGGGCTGATGGTGGACTTG 780
Db      241  AlaLeuLeuLeuAlaGIYGIyGIyHISerArgIleIleGIYProAspGIYGIyAspLeu 260
QY      781  GTGGCGCTCTTGGCGGAAATGAAAGAGGATATCTGTAGCAAACTTGATCCTGGAGTA 840
Db      261  ValaIaIaProlLeuAlaGIuInaenGIuGIYIleLeuYrAlaAsnLeuAspProlGIYVal 280
QY      841  GCGATCTTGTGTAATAATGGCGGAGACCTGTGCTGATCATTAATCCCGTCCGACATTAAT 900
Db      281  ArgIleLeuAlaIyMetAlaIaIaAspProlaGIYHISySerArgProlaIleThr 300
QY      901  CGCTTGCTTAATAGATGGACGCTTAATTAACGGGTAGTGAATTGAAGTGAATCTTGGT 960
Db      301  ArgLeuLeuIleAspArgSerProlYsLeuProValaGIuIleGIuGIyAspLeuArg 320
QY      961  CCTACGCTTTGGGTAAACGCTCTGAGACGGGTGGCAACTCGAAGAAATT 1011
Db      321  ProYrAlaLeuGIYyAlaIaSerGIInTrpGIYAlaGIInaenGIuGIuIle 337
RESULT 5
AD162518
ID      AD162518 standard; protein; 337 AA.
XX      AD162518;
AC      AD162518;
DT      22-APR-2004 (first entry)
DE      Nitriclase polypeptide #193.
XX      XX
KW      Acorvastatin; (R)-ethyl 4-cyano-3-hydroxybutyrate;
KW      (R)-ethyl 4-cyano-3-hydroxybutyric acid; epichlorohydrin;
KW      3-hydroxyglutaronitrile; 3-hydroxyglutaronitrile;
KW      4-cyano-3-hydroxybutyric acid; ethyl-4-cyano-3-hydroxybutyric acid;
KW      mixed hyperlipidaemia; homozygous familial hypercholesterolaemia;
KW      antilipaeamic; enzyme.
OS      Unidentified.
XX      XX
PN      WO2003106415-A2.
XX      XX
PD      24-DEC-2003.
XX      XX
PF      13-JUN-2003; 2003WO-US018840.
XX      XX
PR      13-JUN-2002; 2002US-0389317P.
PR      28-JUN-2002; 2002US-0392944P.
```

```
XX      XX
XX      (DIVE-) DIVERSA CORP.
XX      PA
XX      PT Burk M, Desantis G, Morgan B, Zhu Z;
XX      DR
XX      MPI; 2004-090821/09.
XX      DR N-PSDB; AD162517.
XX      PT Preparation of atorvastatin comprises catalytic conversion of 3-
XX      PT hydroxyglutaronitrile by polypeptide with nitrilase activity, converting
XX      PT obtained 4-cyano-3-hydroxybutyric acid to ethyl-4-cyano-3-hydroxybutyric
XX      PT acid and forming atorvastatin.
PS      Claim 41; SEQ ID NO 386; 253bp; English.
XX      XX
XX      The present invention relates to a method for preparing an atorvastatin
XX      CC intermediate known as (R)-ethyl 4-cyano-3-hydroxybutyrate ((R)-ethyl 4-
XX      CC cyano-3-hydroxybutyric acid). The method comprises optionally converting
XX      CC epichlorohydrin or equivalent to 3-hydroxyglutaronitrile, catalytic
XX      CC conversion of 3-hydroxyglutaronitrile or equivalent to 4-cyano-3-
XX      CC hydroxybutyric acid with a polypeptide having nitrilase activity,
XX      CC converting 4-cyano-3-hydroxybutyric to ethyl-4-cyano-3-hydroxybutyric
XX      CC acid, and converting this to (R)-ethyl 4-cyano-3-hydroxybutyrate. The
XX      CC method involves whole cell processes, cell lysate process, "one pot"
XX      CC processes, and "multi-pot" processes using a variety of parameters.
XX      CC Acorvastatin is used, in conjunction with dietary restriction, in the
XX      CC management of hyperlipidaemia, including hypercholesterolaemia, mixed
XX      CC dyslipidaemia and homozygous familial hypercholesterolaemia. The present
XX      CC sequence represents a nitrilase polypeptide obtained from an
XX      CC environmental sample.
SQ      Sequence 337 AA;
```

## Alignment Scores:

Align. No.:	1,9e-184	Length:	337
Score:	1761.00	Matches:	336
Percent Similarity:	100.0%	Conservative:	1
Best Local Similarity:	99.7%	Mismatches:	0
Query Match:	96.4%	Indels:	0
DB:	8	Gaps:	0

US-09-751-299-3 (1-1014) x AD162518 (1-337)

```
QY      1  ATGAAAGAGCTATCAAGTCCCTGCTGCAAGCCGCCGATTCATGATTTGGAG 60
Db      1  MetIySGIuAlaIleIySValaIaCYsValaGIuAlaIaIaProlIeYrMetAspLeuYs 20
QY      61  GCGAGGTGGAGCAAAACCAATTGATGTGAAGAAAGCAACGTAATAATGCTGCTG 120
Db      21  AlatrIvalAspLYsThrIleGIuLeuMetGIuGIuAlaIaIaArgAsnAlaIaArgLeu 40
QY      121  ATGCGCTTCCGGAACCTTGATTCAGGCTCCCATGGTTCTTGGCTGATCAGCA 180
Db      41  IleAlaPheProlGIuThrTrIleProlGIYrProTrIppheLeuTrIppheAspSerPro 60
QY      181  GCATGGGCAATGCAATTTGTATGCGCAATACCATGAAACTCATTTGAGATTGGCCCT 240
Db      61  AlatrPalameGIInPheValaIaArgIInYrHISGIuAenSerLeuGIuLeuAspGIYPro 80
QY      241  CAAGTTAAGCGGATTTGATGATGACGACCAAGCGTTGGGAATGATGTCACCTGGGGAAG 300
Db      81  GIuAlaIySArgIleSerAspAlaIaIaIySArgLeuGIYIleMetValThrLeuGIYMet 100
QY      301  AGTGAACGGGTGCGGTGACACCTTATCATCATGATGAGTTGATAGCGATATGATGAC 360
Db      101  SerGIuArgValGIyGIYThrLeuYrIleSerGIInTrpPheIleGIYAspAsnGIyAsp 120
QY      361  ACCATTGGGGCCGGGCGAAAGTTGAACCTTATTTGTTGAACGTAATTTGTTGGCGGA 420
Db      121  ThrIleGIYAlaIaArgIArgIySLeuYsPProThrPheValaGIuArgThrLeuPheGIyGIu 140
QY      421  GGGGATGGTTCATGCTAGCGGCTTTTCAGACGCTGTGTGAAGGCTGGCTTAATGC 480
```



Db 141 GlyAspGlySerSerLeuAlaValPheGluThrSerValGlyArgLeuGlyGlyLeuCys 160  
Qy 481 TGTGGGAGCACCTTCAACCGGTACAAATACGCTTTGTATGCAAAATGAAGAGATT 540  
Db 161 CysTrpGluHisLeuGlnProLeuThrLysTrpAlaLeuTrpAlaGlnAsnGluIle 180  
Qy 541 CATTGTGGCGGCTTGGCCGAGCTTTTACCTTTATCCCTAAATGCGGCAAAAGCCCTGGGCGCT 600  
Db 181 HisCysAlaAlaIleTrpProSerPheSerLeuTrpProAlaAlaIleValAlaLeuGlyPro 200  
Qy 601 GATGTCAATGTAGCGGCGCTTGCAGATCTATGCGGTGAAGGCAATGCTCTGTACTACG 660  
Db 201 AspValAsnValAlaAlaSerArgLetyrAlaValGluGlyGlnCysPheValIleuAla 220  
Qy 661 TCGGTGGCGCTTCCATTCATCCATTCGATGCTTTGTACAGATGACGAAAGCAT 720  
Db 221 SerCysAlaLeuValSerGlnSerMetIleAspMetLeuCysThrAspAspGluLysHis 240  
Qy 721 GCGTTGCTTGGCTGGTGGTGGACACTCAGTATCATAGGGCGCTGATGGTGGACTTG 780  
Db 241 AlaLeuLeuLeuAlaGlyGlyGlyHisSerArgLetyrIleGlyProAspGlyGlyAspLeu 260  
Qy 781 GTGCGGCTTCTTCCGAAATATGAGAGGATTTCTTACGCAACCTTGATCTGAGAT 840  
Db 261 ValAlaProLeuAlaGluAsnGluGlyIleLeuTrpAlaAsnLeuAspProGlyVal 280  
Qy 841 CGCATCTTGTGCTAAATGGCGGCAACCTGCTGCTATTTCCCGTCCGACATTTACT 900  
Db 281 ArgIleuAlaLysMetAlaAlaAspProAlaGlyHisTrpSerArgProAspIleThr 300  
Qy 901 CGCTTGCTAATATGATGCGAGCGCTTAATTTACCGGTAGTTGAAATGGAAGTATCTTGT 960  
Db 301 ArgLeuLeuIleAspArgSerProLysLeuProValGluIleGluGlyAspLeuArg 320  
Qy 961 CTTTACGCTTTGGGTAAAGCGCTGTGAGCGGCTGGCAACTCGAAGAAATT 1011  
Db 321 ProTrpAlaLeuGlyLysAlaSerGluThrGlyAlaGlnLeuGluIle 337

RESULT 7  
ADCC23791  
ID ADC23791 standard; protein; 337 AA.  
AC ADC23791;  
XX 18-DEC-2003 (first entry)  
DT Protein sequence (SeqID 58) exhibiting nitrilase activity.  
DE  
DE enzyme; nitrilase; nitrile; cyanohydrin; ammonia; biocatalyst;  
KW enantiomer; chiral medicine.  
XX  
OS unidentified.  
XX  
XX MO2003000840-A2.  
FN  
XX 03-JAN-2003.  
PD  
XX 15-MAY-2002; 2002MO-US015983.  
PE  
XX 21-JUN-2001; 2001US-0300189P.  
PR 30-JUL-2001; 2001US-0309006P.  
PR 22-JAN-2002; 2002US-0351336P.  
XX  
XX (DIVE-) DIVERSA CORP.  
PA (MADP/) MADDEN D.  
XX  
XX Madden M, Desantis G, Chaplin JA, Weiner DP, Milan A, Chi E;  
PI Short JM, Burk M;  
XX WPI; 2003-201417/19.  
DR N-PSDB; ADC23790.  
XX  
XX Novel nitrilase polypeptide, useful for making (R) - or (S) -ethyl-4-cyano-

PT 3-hydroxybutyric acid or (R) - or (S) -mandelic acid or (S) - or (R) -phenyl  
PT lactic acid derivative and for producing pharmaceutical composition, and  
PT food additive.  
PS Claim 40; SEQ ID NO 58; 560pp; English.  
XX  
XX This invention relates to nitrilases and the nucleic acids that encode  
CC these enzymes thereof. Specifically, it refers to polypeptides that  
CC exhibit nitrilase activity, i.e. the ability to directly hydrolyse  
CC nitriles or cyanohydrins into their corresponding carboxylic acids and  
CC ammonia. Nitrilases have commercial utility as biocatalysts for use in  
CC the synthesis of enantiomerically pure aromatic and aliphatic amino  
CC acids, as well as hydroxy acids, which are important for the development  
CC of chiral medicines. Furthermore, the present invention describes  
CC nitrilases, isolated from mesophilic microorganisms, that have improved  
CC activity and stability at increased pH and temperature. They are also  
CC inexpensive, efficient catalysts, have broad substrate specificity and  
CC are capable of chiral differentiation. This polypeptide is a protein  
CC sequence that exhibits nitrilase activity of the invention.  
SQ Sequence 337 AA;  
Alignment Scores:  
Pred. No.: 6,07e-138 Length: 337  
Score: 1339.00 Matches: 252  
Percent Similarity: 84.2% Conservative: 30  
Best Local Similarity: 75.2% Mismatches: 47  
Query Match: 73.3% Indels: 6  
DB: 7 Gaps: 1  
US-09-751-299-3 (1-1014) x ADC23791 (1-337)  
Qy 1 ATGAAAGAGCTATCAAGTGGCTGCGTGCAGCGCCCGGATCTACGATTTGGAG 60  
Db 1 ValLysGluAlaIleLysValAlaCysValGlnAlaAlaProValPheLeuAspLeuAsp 20  
Qy 61 GCGAGGTGGACAAACCATGATGATGGAAGAAGCAGCAGCTAATATGCTGCTG 120  
Db 21 AlatrValAspLysThrValAlaLeuIleGluGlnAlaAlaGlnGlyAlaArgLeu 40  
Qy 121 ATGCGCTTCCGAAACTTGATTCAGGCTTACCCATGCTTCTTGGCTTGACTCACCA 180  
Db 41 IleAlaPheProGluThrTrpIleProGlyTrpProIlePheLeuTrpLeuAspSerPro 60  
Qy 181 GCATGGGCAATGCAATTTGTACGCCAATACATGAGAACTCATTTGAGTGGATGGCCT 240  
Db 61 AlatrGlyMetGlnPheValArgArgTrpHisGluAsnSerLeuValLeuAspSerPro 80  
Qy 241 CAAGCTAAGGCAATTCAGATGAGCGCAAGGCGTTGGGAATCATGTCACCCCTGGGATG 300  
Db 81 GlnAlaLysArgLysSerGluAlaAlaGlnArgAlaGlyIleTrpValAlaLeuGlyTrp 100  
Qy 301 AGTGAACGGGTCCGCTGAGCACCTTTACATCAGTGGTTCATAGGCGATTAATGTGAC 360  
Db 101 SerGluArgValSerGlyThrLeuTrpMetGlyGlnTrpLeuIleAspAspLysGlyGlu 120  
Qy 361 ACATTTGGGCGCCCGGCAAAAGTTGAACCTTCTTTGTAACGTACTTTGTTGGCGAA 420  
Db 121 ThrAlaGlyLeuArgArgLysLeuLysProThrHisValGlnArgThrLeuPheGlyGlu 140  
Qy 421 GGGAGTGGTTCATGCTAGCGGTGTTGACGCTGTGTGGAAGGCTGGGTGGCTTATGC 480  
Db 141 GlyAspGlySerSerLeuSerThrPheAspThrProLeuGlyValLeuGlyGlyLeuCys 160  
Qy 481 TGTGGGAGCACCTTCAACCGGTACAAATACGCTTTGTATGCAAAATGAAGAGATT 540  
Db 161 CysTrpGluHisLeuGlnProLeuSerLysTrpAlaLeuTrpAlaGlnAsnGluIle 180  
Qy 541 CATTGTGGCGGCTTGGCCGAGCTTTTACCTTTATCCCTAAATGCGGCAAAAGCCCTGGGCGCT 600  
Db 181 HisPheAlaAlaIleTrpProSerPheSerLetyrArgGlnAlaThrGlnValLeuGlyPro 200  
Qy 601 GATGTCAATGTAGCGGCGCTTGCAGATCTATGCGGTGAAGGCAATGCTCTGTACTACG 660

DB 201 GluValAsnValAlaIaIaSerArgIleTyrAlaValGluGlyIncysPheValIleuAla 220  
QY 661 TCGTGGCGCGCTTCGTTGCATCATCATGATGCTTGTGTAGAGATGAGAAAGCAT 720  
DB 221 SerCysAlaIleuValSerProGluMetIleGluMetLeuCysTrpAspGluSerLysHis 240  
QY 721 GCGTTGCTTGGCTGGTGTGACACTGATCATAGAGGCGCTGATGTGACTTGG 780  
DB 241 SerIleuGlnIaIaGlyGlyIleTyrSerArgIleIleGlyProAspGlySerAspIleu 260  
QY 781 GTGCGCGCTCTTGGCGAAATGAAAGAGGATTTCTTACGAAACCTTGATCTGGAGTA 840  
DB 261 AlaArgProIeuGlyGluAsnGluGlyIleLeuTyrAlaThrIleuAspProIaAla 280  
QY 841 CGCATCGCTGCTAAATAGCGGAGACCTGCTGATCATATTCCTCCGCTCCGACATTACT 900  
DB 281 ArgIleTyrAlaIleTyrThrAlaIaAspProIaIaGlyIleTyrSerArgProAspValThr 300  
QY 901 CGCTTGCTAATAGATCGAGCGCTTAAATTCACCGGTAGTTGAAATGAAAGTGATCTTCT 960  
DB 301 ArgIleuIleuIleAsnArgSerAlaAsnGlnProValIaIaGluValGlyArgGluIle-- 319  
QY 961 CCTTACCGCTTGGGTAAAGCGCTGAGACGCGGTGCGAACCTCGAA 1005  
DB 320 -----ProIaSerAlaGlnGlyPheGluValGln 329  
RESULT 8  
ADC23781  
ID ADC23781 standard; protein; 337 AA.  
XX  
AC ADC23781;  
XX  
DT 18-DEC-2003 (first entry)  
XX  
DE Protein sequence (SeqID 48) exhibiting nitrilase activity.  
XX  
KM enzyme; nitrilase; nitrile; cyanohydrin; ammonia; biocatalyst;  
XX  
XX enantiomer; chiral medicine.  
XX  
OS Undefined.  
XX  
PN W02003000840-A2.  
XX  
PD 03-JAN-2003.  
XX  
PF 15-MAY-2002; 2002MO-US015983.  
XX  
PR 21-JUN-2001; 2001US-0300189P.  
XX  
PR 30-JUL-2001; 2001US-030906P.  
XX  
PR 22-JAN-2002; 2002US-0351336P.  
XX  
PA (DIVE-) DIVERSA CORP.  
XX  
PA (MAD/) MADDER D.  
XX  
PI Madden M, Desantis G, Chaplin JA, Weiner DP, Milan A, Chi E;  
PI Short JM, Burk M;  
XX  
DR WPI: 2003-201417/19.  
XX  
DR N-PSDB; ADC23780.  
XX  
XX  
PT Novel nitrilase polypeptide, useful for making (R)- or (S)-ethyl-4-cyano-  
PT 3-hydroxybutyric acid or (R)- or (S)-mandelic acid or (S)- or (R)-phenyl  
PT lactic acid derivative and for producing pharmaceutical composition, and  
PT food additive.  
XX  
XX  
PS Claim 40; SEQ ID NO 48; 560bp; English.  
XX  
XX  
CC This invention relates to nitrilases and the nucleic acids that encode  
CC these enzymes thereof. Specifically, it refers to polypeptides that  
CC exhibit nitrilase activity, i.e. the ability to directly hydrolyse  
CC nitriles or cyanohydrins into their corresponding carboxylic acids and  
CC ammonia. Nitrilases have commercial utility as biocatalysts for use in

CC the synthesis of enantiomerically pure aromatic and aliphatic amino  
CC acids, as well as hydroxy acids, which are important for the development  
CC of chiral medicines. Furthermore, the present invention describes  
CC nitrilases, isolated from mesophilic microorganisms, that have improved  
CC activity and stability at increased pH and temperature. They are also  
CC inexpensive, efficient catalysts, have broad substrate specificity and  
CC are capable of chiral differentiation. This polypeptide is a protein  
CC sequence that exhibits nitrilase activity of the invention.  
XX  
SQ Sequence 337 AA;

Alignment Scores:  
Pred. No.: 6.07e-138 Length: 337  
Score: 1339.00 Matches: 252  
Percent Similarity: 84.2% Conservative: 30  
Best Local Similarity: 75.2% Mismatches: 47  
Query Match: 73.3% Indels: 6  
DB: Gaps: 1

US-09-751-299-3 (1-1014) x ADC23781 (1-337)

QY 1 ATGAAAGAGCTATCAAGTTCGCTGCGCAAGCCGCCGATCTACATGATTTGGAG 60  
DB 1 ValysGluIaIaIeIysValAlaCysValGlnIaIaIaProValPheIeAspIeAsp 20  
QY 61 GCGACGCTGGACAAACCATTTGATGATGAAAGACAGACAGCTAATATGCTCGCTG 120  
DB 21 AlaThrValAspIysThrIaIaIaLeuIleGluIuIaIaIaAsnGlyAlaArgIleu 40  
QY 121 ATCGCTTTCGGAACCTTGATTCAGAGCTACCCATGCTTCTTGGCTGACTCAACA 180  
DB 41 IleAlaPheProGluIuIuIuIleProGlyTyrProIleuIuIuIeAspSerPro 60  
QY 181 GCATGGCAATGCATTTGTACGCCAATACCATAGAACTCATTTGAGATTGGCCCT 240  
DB 61 AlaIlePylMetGlnPheValAlaArgIleTyrHisGluIuIuIeAspSerPro 80  
QY 241 CAAGCTAAGCGCATTTGAGATGACGACGAGCGGTGGAAATCATAGTACCCCTGGGATG 300  
DB 81 GlnIaIaIysArgIleSerGluIaIaIaGlnArgIaIaGlyIleTyrValAlaIeGlyTyr 100  
QY 301 AGTGAACGGCTGGTGGACACCTTTACATCATGATGCTTGCATAGCGCATATNGTGAC 360  
DB 101 SerGluIuIuIeIleSerGlyIuIuIeIleTyrIleuTyrIleuIleAspIleGlyGlu 120  
QY 361 ACCATGGGGCGCGGCAAGTTGAAACCTACTTTGTGTGAACGATCTTGTGGCGAA 420  
DB 121 ThrIaIeIuIeIuIa 140  
QY 421 GGGATGATTCATGCTAGACGCTTTCGAGACGCTGTGGAGGCTGGGTGCTTATGC 480  
DB 141 GlyAspGlySerSerIleuSerThrPheAspIleuIleuIleuIleuIleuIleuIleu 160  
QY 481 TGTGGAGACCTTCAACCGCTTCAACAAAGACCTTGTATGACAAAGAGAGATT 540  
DB 161 CysIlePylIuIuIeIuIuIeIuIeIuIeIuIeIuIeIuIeIuIeIuIeIuIeIuIe 180  
QY 541 CATGTGCGGCTTGGCGAGCTTATGCTTATCTATGCGGCGAAAGCCCTGGGGCT 600  
DB 181 HisPheAlaIa 200  
QY 601 GATGCTAATGAGCGGCTCTCGAATCTATGCGGTGAAAGGCAATGTTGTACTAGCG 660  
DB 201 GluValAsnValAlaIaIaSerArgIleTyrAlaValGluGlyIncysPheValIleuAla 220  
QY 661 TCGTGGCGCGCTTCGTTGCATCATCATGATGCTTGTGTAGAGATGAGAAAGCAT 720  
DB 221 SerCysAlaIleuValSerProGluMetIleGluMetLeuCysTrpAspGluSerLysHis 240  
QY 721 GCGTTGCTTGGCTGGTGTGACACTGATCATAGAGGCGCTGATGTGACTTGG 780  
DB 241 SerIleuGlnIaIaGlyGlyIleTyrSerArgIleIleGlyProAspGlySerAspIleu 260







XX 15-MAY-2003; 2003WO-US015712.  
 PF  
 XX  
 PR 15-MAY-2002; 2002US-00146772.  
 PR 09-SEP-2002; 2002US-00241742.  
 XX  
 PA (DIVE-) DIVERSA CORP.  
 XX  
 PI Desantis G, Short JM, Burk M, Wong K, Farwell R, Chatman K;  
 XX WPI; 2004-090637/09.  
 DR N-PSDB; ADG93582.  
 DR  
 XX  
 PT New isolated or recombinant nucleic acid encoding a polypeptide having  
 PT nitrilase activity, useful for screening enantioselective transformation.  
 XX  
 PS Claim 44; SEQ ID NO 48; 295pp; English.  
 XX  
 CC This invention is related to a novel isolated or recombinant nucleic acid  
 CC encoding a protein having nitrilase activity. Nitrilase's are capable of  
 CC converting nitrile's directly to carboxylic acids and have great  
 CC potential for use in industrial chemical processes. The isolated  
 CC nitrilase proteins of the invention have increased activity and stability  
 CC at increased pH and temperature when compared to those conventionally  
 CC used. In addition, the nucleic acid of the invention is useful for  
 CC screening enantioselective transformation. The present sequence is that  
 CC of a nitrilase enzyme of the invention.  
 XX  
 XX Sequence 337 AA;  
 SQ  
 Alignment Scores:  
 Pred. No.: 6.07e-138 Length: 337  
 Score: 1339.00 Matches: 252  
 Percent Similarity: 84.2% Conservative: 30  
 Best Local Similarity: 75.2% Mismatches: 47  
 Query Match: 73.3% Indels: 6  
 DB: 8 Gaps: 1  
 US-09-751-299-3 (1-1014) x ADG93583 (1-337)

DB 161 CyetrrpGluHsleuGlnProleuSerlyrAlaleuTrAlaGlnAenGluGluLe 180  
 QY 541 CATTCGCGGCTTGGCCGAGCTTTAGCCTTATCTTAATGGGCGAAAGCCCTGGGCGCT 600  
 DB 181 HlsheAlaAlatrrProserPheSerlleTyrrglnAlatrrGlnuAlleuGlyPro 200  
 QY 601 GATGTCAATGTAGGGGCTCCGAATCTATGCGGTGAGGCGCAATGCTTCTGCGG 660  
 DB 201 GluValaenValAlaAlaSerArgIleTrAlaValaGluGlyGlnCysPheValLeuAla 220  
 QY 661 TCGTGCGGCTCGTTTCAACATCCATGATGATATGCTTTGTACAGATGACGAAGCAT 720  
 DB 221 SerCysAlaLeuValSerProGluMetIleGluMetLeuCystrhrPabGluSerlyHis 240  
 QY 721 GCGTTCCTTGGCTGGTGTGACACTCACGATATCAATAGGCGCTGATGTGTGACTTG 780  
 DB 241 SerLeuLeuGlnAlaGlyGlyTyrrSerArgIlelleGlyProAapGlySerAapLeu 260  
 QY 781 GTGCGGCTTGGCCGAATGAAAGGGTATTCCTTACCGCAACCTTGATCCTGGAGTA 840  
 DB 261 AlaArgProleuGlyGluAenGluGlyIleLeuTrAlaThrLeuAapProAlaAla 280  
 QY 841 CGCATCCTTGTAAATGGCGGAGACCTGTGTGATTAATTCGCGCGACATTAAT 900  
 DB 281 ArgIleTrAlaAlaYstrhrAlaAlaAapProAlaGlyHisTyrrSerArgProAapValThr 300  
 QY 901 CGCTTGCTAATGATGCGACCCCTTAATTAACGGTATGTAATGAAAGGTATCTTGT 960  
 DB 301 ArgLeuLeuIleAenAargSerAlaAenGlnProValAlaGluValAlaGlyArgGluIle 319  
 QY 961 CCTTACGCTTGGGTAAAGGCTCTGAGACGGGTGCGGCAACTCGAA 1005  
 DB 320 -----ProAlaSerAlaGlnGlyPheGluValGlu 329  
 RESULT 12  
 ADG93593  
 ID ADG93593 standard; protein; 337 AA.  
 XX  
 AC ADG93593;  
 XX  
 DT 11-MAR-2004 (first entry)  
 XX  
 DE Nitrilase enzyme amino acid sequence SeqID58.  
 XX  
 KM nitrilase; nitrile; carboxylic acid; chemical process; pH; temperature;  
 XX enantioselective transformation; enzyme.  
 OS Unidentified.  
 XX  
 PN WO2003097810-A2.  
 PD 27-NOV-2003.  
 XX  
 PF 15-MAY-2003; 2003WO-US015712.  
 XX  
 PR 15-MAY-2002; 2002US-00146772.  
 PR 09-SEP-2002; 2002US-00241742.  
 XX  
 XX (DIVE-) DIVERSA CORP.  
 PA  
 XX Desantis G, Short JM, Burk M, Wong K, Farwell R, Chatman K;  
 XX WPI; 2004-090637/09.  
 DR N-PSDB; ADG93592.  
 DR  
 XX  
 PT New isolated or recombinant nucleic acid encoding a polypeptide having  
 PT nitrilase activity, useful for screening enantioselective transformation.  
 XX  
 PS Claim 44; SEQ ID NO 58; 295pp; English.  
 XX  
 CC This invention is related to a novel isolated or recombinant nucleic acid  
 CC encoding a protein having nitrilase activity. Nitrilase's are capable of



SQ Sequence 337 AA;

## Alignment Scores:

Pred. No.: 6.07e-138 Length: 337  
 Score: 1339.00 Matches: 252  
 Percent Similarity: 84.2% Conservative: 30  
 Best local Similarity: 75.2% Mismatches: 47  
 Query Match: 73.3% Indels: 6  
 DB: 8 Gaps: 1

US-09-751-299-3 (1-1014) x ADI62180 (1-337)

```

QY 1 ATGAAAGAGCTATCAAGTGGCGCTGCAAGCCGCCGATCATAGATTGGAG 60
DB 1 Vallysgluatlelyvalalacysvalglnalalaprovalpheleuasp 20
QY 61 GCGAGGTGAGCAAAACCATTTGATGAGAGAGAGACGTAATATGCTGCTG 120
DB 21 AlathrvalasplyrthValAlaleuilegluglnalalargasnclAlalargleu 40
QY 121 ATGCGCTTCCGAAACTTGATTCAGGCTACCCATGGTTCTTGGCTGACTACCA 180
DB 41 llaalphepogluThrTpIleProgllyrProItrpHeleutrpLeuAspserPro 60
QY 181 GCATGGGCAATGCAATTTGTACGCCAATACATGAGACTCATGAGATTGGAGCCCT 240
DB 61 AlatrpglmetGlnPheValArgtryrhsgluasnSerLeuValLeuAspserPro 80
QY 241 CAAGCTAAGCGCATTTTCAGATGACGCCAAGCGTTGGGAAATCATGTCACCCCTGGGATG 300
DB 81 GlnalalysarglieserGlnalalaglnargalaglylleTyValAlaleucllyr 100
QY 301 AGTGAACGGGTGGTGACCCCTTTACATCAGTCAGTGGTTCAATAGCGATPATGTGAC 360
DB 101 SergluargValserGlyThrleuTyrmwetGlyIntPleuileAspAsplysglylu 120
QY 361 ACCATTGGGGCCCGCGGAAATTTGAACCTTTGTTGAACGTAATTTGTTGGCGAA 420
DB 121 ThrIaglyleuargarglysleuysProThrhisValGluargThleuphelelylu 140
QY 421 GGGAGATGGTTATCGCTAGCGGTTTTCAGACGTCGTTGGAAGGCTGGGTGCTTATGC 480
DB 141 GlyaspgllyserseleuSerThrPheasphtrProleuGlyValleuclglyleuys 160
QY 481 TGTGGAGACCTTCAACCGCTAACAAATACGCTTGTATGCAAAATGAAGAGATT 540
DB 161 CysItrpgluhisleuGlnProleuSerlyrAlaleuTyrlaGlnAsngluclulle 180
QY 541 CATGTGGGCTTGGCGGACGCTTATGCTTATCTTAATGGGCGAAACCCCTGGGCT 600
DB 181 HispheaialatrpProserPheSerileTyraGlnalathrGluValleucllyPro 200
QY 601 GATGCAATGAGGGCGCTTCGATCATGCGGTTGAAGGCAATGCTGCTACTACG 660
DB 201 GlnValaenValalalaserarglleTyrlalavalaglnGlyGlnCypheValleucl 220
QY 661 TCGTGTGGCTCGTTTCACAATCCATGATGCATATGCTTGTATGACATGAGAAAGCAT 720
DB 221 SerCysalaleuValserProgluWetIlegluWetleuCystrhaSpgluSerlyshis 240
QY 721 GCGTTGCTTGGCTGGGTGGTGGACATCAGTATCATAGGGCCCTGAGTGGTGGCTTG 780
DB 241 SerleuLeuGlnalaglylyGlylyrzerarglleleleGlyProAsplyserAspleu 260
QY 781 GTCGGGCTCTTGGCGAAATGAAGAGGTAATTTCTACGCAAACTTGATCTGAGATA 840
DB 261 AlaargProleuGlyGlnAsnglucllylleuTyrlalathrLeuAspProalalala 280
QY 841 CGCATCTTGTCTAAATGCGCGACAGCCCTGCTGCTATTTCCGTCGCGACATTAAT 900
DB 281 ArgIleTyrlalalyThrAlalalaspProalaglyhisTyserArgProAspValthr 300
QY 901 CGCTTGCTAATATGATCGACGCCCTTAATTCACCGGTAAGTTGAATTTGAAGGATCTTGGT 960
  
```

DB 301 ArgleuLeuileanAargserAlaasnGlnProValValGluValGlyArgGluLeu--- 319  
 QY 961 CTTTACGCTTGGGTAAAGCTCTGAGACGGGTGGCGCACTTCGAA 1005  
 DB 320 -----ProAlaserAlaglncllyPheGluValGlu 329

## RESULT 14

ADI62190

ADI62190 standard; protein, 337 AA.

AC ADI62190;

DT 22-APR-2004 (first entry)

DE Nitriase polypeptide #29.

KW Acorvastatin; (R)-ethyl 4-cyano-3-hydroxybutyrate;  
 (R)-ethyl 4-cyano-3-hydroxybutyric acid; epichlorohydrin;  
 3-hydroxyglutaronitrile; 3-hydroxyglutaronitrile;  
 4-cyano-3-hydroxybutyric acid; ethyl 4-cyano-3-hydroxybutyric acid;  
 mixed hyperlipidaemia; homozygous familial hypercholesterolaemia;  
 antilipemic; enzyme.

KW Unidentified.

XX WO2003106415-A2.

XX 24-DEC-2003.

XX 13-JUN-2003; 2003WO-US018840.

XX 13-JUN-2002; 2002US-0389317P.

XX 28-JUN-2002; 2002US-0392944P.

XX (DIVE-) DIVERSA CORP.

XX Burk M, Desantis G, Morgan B, Zhu Z;

XX WP1; 2004-090821/09.

XX N-PSDB; ADI62189.

XX Preparation of atorvastatin comprises catalytic conversion of 3-

PT hydroxyglutaronitrile by polypeptide with nitriase activity, converting  
 PT obtained 4-cyano-3-hydroxybutyric acid to ethyl-4-cyano-3-hydroxybutyric  
 PT acid and forming atorvastatin.

XX Claim 41; SEQ ID NO 58; 253bp; English.

XX The present invention relates to a method for preparing an atorvastatin

CC intermediate known as (R)-ethyl 4-cyano-3-hydroxybutyrate ((R)-ethyl 4-  
 CC cyano-3-hydroxybutyric acid). The method comprises optionally converting  
 CC epichlorohydrin or equivalent to 3-hydroxyglutaronitrile, catalytic  
 CC conversion of 3-hydroxyglutaronitrile or equivalent to 4-cyano-3-  
 CC hydroxybutyric acid with a polypeptide having nitriase activity,  
 CC converting 4-cyano-3-hydroxybutyric to ethyl-4-cyano-3-hydroxybutyric  
 CC acid, and converting this to (R)-ethyl 4-cyano-3-hydroxybutyrate. The  
 CC method involves whole cell processes, cell lysate processes, "one pot"  
 CC processes, and "multi-pot" processes using a variety of parameters.  
 CC Atorvastatin is used, in conjunction with dietary restriction, in the  
 CC management of hyperlipidaemia, including hypercholesterolaemia, mixed  
 CC dyslipidaemia and homozygous familial hypercholesterolaemia. The present  
 CC sequence represents a nitriase polypeptide obtained from an  
 CC environmental sample.

SQ Sequence 337 AA;

## Alignment Scores:

Pred. No.: 6.07e-138 Length: 337  
 Score: 1339.00 Matches: 252  
 Percent Similarity: 84.2% Conservative: 30  
 Best local Similarity: 75.2% Mismatches: 47  
 Query Match: 73.3% Indels: 6



DB: 8 Gaps: 1

US-09-751-299-3 (1-1014) x ADI64311 (1-337)

Search completed: April 27, 2006, 01:18:29  
Job time : 177.493 secs

```
QY      1 ATGAAAGAGCTATCAAGGTGCGCTTGCGTCAAGCCGCCGATCTACATGATTTGGAG 60
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db      1 VallyeGluAlaIleValValAlaCyValGlnAlaIleProValPheLeuSerPro 20
QY      61 GGCACGCTGGACAAACCATTTGATGAGAAAGCAGACGTAATATGCTGCTG 120
      ||::::::::::::::::::::::::::::::::::::::::::::::::::::
Db      21 AlaThrValAspLysThrValAlaLeuIleGlnGlnAlaIleArgAsnGlyAlaArgLeu 40
QY      121 ATGCGCTTCCGGAACCTTGATTCAGGCTACCCATGGTTCTTGCTTGACTCACA 180
      ||::::::::::::::::::::::::::::::::::::::::::::::::::::
Db      41 IleAlaPheProGluThrTrpIleProGlyTyrProTrpPheLeuTrpLeuAspSerPro 60
QY      181 GCATGGGCAATGCATTTTGTACGCCAATACATGAGAACTCATTTGAGTTGATGGCCCT 240
      ||::::::::::::::::::::::::::::::::::::::::::::::::::::
Db      61 AlaTrpGlyMetGlnPheValArgArgTyrHisGlnAsnSerLeuValLeuAspSerPro 80
QY      241 CAAAGCTAACCGCATTTCAATGACAGCAAGCGGTTGGAAATCATGTGCACCTGGGGATG 300
      ||::::::::::::::::::::::::::::::::::::::::::::::::::::
Db      81 GlnAlaLysArgIleSerGlnAlaIleArgAlaGlyIleTyrValAlaLeuGlyTyr 100
QY      301 AGTGAACGGGTGCGTGCACCCCTTTCATCAGTCAGTGGTTCATAGCGCATATGGTGAC 360
      ||::::::::::::::::::::::::::::::::::::::::::::::::::::
Db      101 SerGlnArgValSerGlyThrLeuTyrMetGlyGlnTrpLeuIleAspAspLysGlyGln 120
QY      361 ACCATTGGGGCCCGCGGAAGTTGAAACCTACTTTTGTGAAAGTACTTTGTTCCGGCA 420
      ||::::::::::::::::::::::::::::::::::::::::::::::::::::
Db      121 ThrAlaGlyLeuAlaArgLysLeuLysProThrHisValGlnArgThrLeuPheGlyGln 140
QY      421 GGGGATGGTTCATCCGCTTACCGGTTTCGAAAGCTGTGAAAGCTGGGCTTATGC 480
      ||::::::::::::::::::::::::::::::::::::::::::::::::::::
Db      141 GlyAspGlySerSerLeuSerThrPheAspThrProLeuGlyValLeuGlyLeuCyS 160
QY      481 TGTGGGAGACCTTCAACCGCTAACAAATACGCTTGTATGACAAATAGAGAGATT 540
      ||::::::::::::::::::::::::::::::::::::::::::::::::::::
Db      161 CysTrpGlnHisLeuGlnProLeuSerLysTyrAlaLeuTyrAlaGlnAsnGlnGluIle 180
QY      541 CATTTGCGCGCTTGGCGGAGCTTTAGCCTTTATCCCTAATGCGGCAAGCCCTGGGGCCT 600
      ||::::::::::::::::::::::::::::::::::::::::::::::::::::
Db      181 HisPheAlaIleAlaTrpProSerPheSerIleTyrArgGlnAlaThrGluValLeuGlyPro 200
QY      601 GATGTCAATGTAGCGCGCTCTGGAATCTATGCCGTTGAAGGGCAATGCTTCGTACTAGCG 660
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db      201 GluValAsnValAlaIleAspArgIleTyrAlaValAlaGlnGlyGlnCysPheValLeuAla 220
QY      661 TGTGTGGCGCTGTTTCAATCCATGCATGCATATGCTTGTATGACAAACCTTGATCCTGGAGTA 720
      ||::::::::::::::::::::::::::::::::::::::::::::::::::::
Db      221 SerCyAlaLeuValSerProGluMetIleGlnMetLeuCyThrAspGlnSerLysHis 240
QY      721 GCGTTCCTTCTGAGCTGTGTGAGACACTACGATATCAGAGGCTGATGTGACTTG 780
      ::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db      241 SerLeuLeuGlnAlaGlyGlyGlyTyrSerArgIleIleGlyProAspLysSerAspLeu 260
QY      781 GTGCGGCTCTTTCGCGAAATGAAGGGTATTCTCTACGCAACCTTGATCCTGGAGTA 840
      ||::::::::::::::::::::::::::::::::::::::::::::::::::::
Db      261 AlaArgProLeuGlyGlnAsnGlnGlnGlyIleLeuTyrAlaThrLeuAspProAlaAla 280
QY      841 CGCATCCCTTGTAAATGGCGGACAGCCCTGCTGTCATTAATCCGTCGCCGACATTA 900
      ||::::::::::::::::::::::::::::::::::::::::::::::::::::
Db      281 ArgIleTyrAlaLysThrAlaIleAspProAlaGlnLysTyrSerArgProAspValThr 300
QY      901 CCGTTCCTATAGATGACAGCCCTAAATACCGGTAGTGAATGAAGTATGATCTTCGT 960
      ||::::::::::::::::::::::::::::::::::::::::::::::::::::
Db      301 ArgLeuLeuIleAsnArgSerAlaAsnGlnProValValGlnValGlyArgGluIle---- 319
QY      961 CCTTACGCTTTGGGTAAACGCTGTGAGACGGGTGCGCAACTCGAA 1005
      ||::::::::::::::::::::::::::::::::::::::::::::::::::::
Db      320 -----ProIleSerAlaGlnGlyPheGlnValGln 329
```

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OW nucleic - protein search, using frame\_plus\_n2p model

Run on: April 27, 2006, 01:35:46 ; Search time 7.1054 Seconds  
(without alignments)  
2359.702 Million cell updates/sec

Title: US-09-751-299-3

Perfect score: 1826

Sequence: 1 atgaagaagatcatcaaggt.....cgcgaactcgaagaatttga 1014

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 1144120

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODE=frame+ n2p.model -DEV=xip  
-O=/abs/ABSSWEB.spool/US09751299/runat.26042006.090039.18719/app.query.fasta\_1  
-DB=Issued\_Patents\_AA -QFMT=fasta -SUFFIX=rai -MINMATCH=0.1 -LOOCL=0  
-IOOEXT=0 -UNITS=bits -START=1 -END=1 -MATRIX=blosum62 -TRANS=human40.cdi  
-LIST=45 -DOCLIGN=200 -THR\_SCORE=pct -THR\_MAX=100 -THR\_MIN=0 -ALIGN=15  
-MODE=LOCAL -OUTFMT=pico -NORM=ext -HEADSIZE=500 -MINLEN=0 -MAXLEN=2000000000  
-HOST=abss02p -USER=US09751299 @CGN 1 1 101 @runat.26042006.090039.18719  
-NCPU=6 -ICPU=3 -NO\_MMAP -NEG\_SCORES=0 -WAIT -DSPLOCK=100 -LONGLOG  
-DEV\_TIMEOUT=120 -WARN\_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6  
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Issued\_Patents\_AA:\*

1: /cgn2\_6/ptodata/1/1aa/5\_COMB.pep:\*  
2: /cgn2\_6/ptodata/1/1aa/6\_COMB.pep:\*  
3: /cgn2\_6/ptodata/1/1aa/H\_COMB.pep:\*  
4: /cgn2\_6/ptodata/1/1aa/PCTUS\_COMB.pep:\*  
5: /cgn2\_6/ptodata/1/1aa/RE\_COMB.pep:\*  
6: /cgn2\_6/ptodata/1/1aa/Backfile1.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	955.5	52.3	344	1	US-08-690-493-1 Sequence 1, Appli
2	822	45.0	356	2	US-09-806-876A-2 Sequence 2, Appli
3	776	42.5	369	2	US-09-823-373-5 Sequence 5, Appli
4	776	42.5	369	2	US-09-823-373-14 Sequence 14, Appli
5	638	34.9	354	1	US-08-447-702-5 Sequence 5, Appli
6	638	34.9	354	1	US-08-465-615-5 Sequence 5, Appli
7	152	8.3	267	2	US-09-902-540-14601 Sequence 14601, A
8	139.5	7.6	315	2	US-10-105-2948-2 Sequence 2, Appli
9	139	7.6	311	1	US-07-917-111-5 Sequence 5, Appli
10	139	7.6	311	1	US-07-917-111-6 Sequence 6, Appli
11	139	7.6	311	1	US-08-479-638-5 Sequence 5, Appli
12	139	7.6	311	1	US-08-479-638-6 Sequence 6, Appli

13	139	7.6	311	1	US-08-294-871A-70 Sequence 70, Appli
14	139	7.6	311	2	US-08-876-398A-70 Sequence 40, Appli
15	133	7.3	303	1	US-08-294-871A-46 Sequence 46, Appli
16	133	7.3	303	2	US-08-876-398A-46 Sequence 46, Appli
17	132	7.2	303	1	US-08-294-871A-36 Sequence 36, Appli
18	132	7.2	303	2	US-08-876-398A-36 Sequence 36, Appli
19	131	7.2	303	1	US-08-294-871A-16 Sequence 16, Appli
20	131	7.2	303	1	US-08-294-871A-16 Sequence 16, Appli
21	131	7.2	303	1	US-08-294-871A-48 Sequence 48, Appli
22	131	7.2	303	1	US-08-294-871A-48 Sequence 48, Appli
23	131	7.2	303	1	US-08-294-871A-58 Sequence 58, Appli
24	131	7.2	303	1	US-08-294-871A-68 Sequence 68, Appli
25	131	7.2	303	2	US-08-876-398A-4 Sequence 4, Appli
26	131	7.2	303	2	US-08-876-398A-16 Sequence 16, Appli
27	131	7.2	303	2	US-08-876-398A-34 Sequence 34, Appli
28	131	7.2	303	2	US-08-876-398A-48 Sequence 48, Appli
29	131	7.2	303	2	US-08-876-398A-58 Sequence 58, Appli
30	131	7.2	303	2	US-08-876-398A-68 Sequence 68, Appli
31	130	7.1	303	1	US-08-294-871A-38 Sequence 38, Appli
32	130	7.1	303	1	US-08-294-871A-44 Sequence 44, Appli
33	130	7.1	303	1	US-08-294-871A-50 Sequence 50, Appli
34	130	7.1	303	1	US-08-294-871A-56 Sequence 56, Appli
35	130	7.1	303	1	US-08-294-871A-60 Sequence 60, Appli
36	130	7.1	303	2	US-08-876-398A-38 Sequence 38, Appli
37	130	7.1	303	2	US-08-876-398A-44 Sequence 44, Appli
38	130	7.1	303	2	US-08-876-398A-50 Sequence 50, Appli
39	130	7.1	303	2	US-08-876-398A-56 Sequence 56, Appli
40	130	7.1	303	2	US-08-876-398A-60 Sequence 60, Appli
41	130	7.1	304	1	US-08-900-711-5 Sequence 5, Appli
42	130	7.1	304	1	US-08-815-356-2 Sequence 2, Appli
43	130	7.1	304	1	US-08-415-343B-5 Sequence 5, Appli
44	130	7.1	304	1	US-08-762-433-1 Sequence 1, Appli
45	130	7.1	304	1	US-09-001-219-1 Sequence 1, Appli

#### ALIGNMENTS

RESULT 1  
US-08-690-493-1  
; Sequence 1, Application US/08690493  
; Patent No. 5872000  
; GENERAL INFORMATION:  
; APPLICANT: Yu, Fujio  
; TITLE OF INVENTION: No. 5872000el Nitrlaee Gene  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Steindberg, Raskin & Davidson, P.C.  
; STREET: 1140 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 10036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: MS-DOS Editor  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/690,493  
; FILING DATE: 31 JUL-1996  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: UP 213061/1995  
; FILING DATE: 31-JUL-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Davidson, Clifford M  
; REGISTRATION NUMBER: 32,728  
; REFERENCE/DOCKET NUMBER: 3821005  
; TELEPHONE: (212) 768-3800  
; TELEFAX: (212) 362-2124  
; INFORMATION FOR SEQ ID NO: 1:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 344 amino acid residues

```

; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; ORGANISM: Gordona terrae
; STRAIN: MA-1
; CELL TYPE: unicellular organism
; FEATURE INFORMATION: Xaa is Met or a deletion
; OTHER INFORMATION: Xaa is Met or a deletion
; US-08-690-493-1
```

## Alignment Scores:

Pred. No.:	7,986-97	Length:	344
Score:	955.50	Matches:	182
Percent Similarity:	73.48	Conservative:	47
Best Local Similarity:	58.38	Mismatches:	82
Query Match:	52.34	Indels:	1
DB:	1	Gaps:	1

US-09-751-299-3 (1-1014) x US-08-690-493-1 (1-344)

```

QY 16 AAGTCGCTGCTGCAAGCCGCCGATCTACATGATTTGAGCGGACGATGACAA 75
    |||||
DB 10 LysalaalaValThrGlnalaGlnProValTrpPheAspLeuSerAlaThrValAspLys 29
    |||||
QY 76 ACCATTGATGATGGAAGAAGACAGACATTAATATGCTCGCTGATCGCTTCCGGA 135
    |||||
DB 30 ThrilealaValGlnGlnalaSerAlaGlyAlaAspLeuIleAlaPheProGln 49
    |||||
QY 136 ACTTGATTCAGGCTTACCCATGATGTTCTTTGCTTGACTCCAGCATGGGCAATGCA 195
    |||||
DB 50 ThrTrpIleProGlnLysTrpTrpPheLeuTrpLeuAspSerValAlaTrpGlnSerGln 69
    |||||
QY 196 TTTTGACCGCATACATGAGAACATTCATGAGTTGAGTGGCCGCAAGCTTAAGCGCAT 255
    |||||
DB 70 TyrPheIleArgLysTrpGlnAsnSerLeuAspLeuAspGlySerGlnPheAlaIle 89
    |||||
QY 256 TCAGATGAGCCAGCAAGCTTGGGAAATCATGTGACCTTGGGGATGAGAACGGTCCGT 315
    |||||
DB 90 ArgGlnAlaAlaArgLysAsnAspIleAlaIleThrMetGlyPheSerGlnArgGlyHis 109
    |||||
QY 316 GGCACCCCTTACATCAGTCAGTCATGAGCGCATATGATGACACCATTTGGGGCCGG 375
    |||||
DB 110 GlySerLeuTyrMetGlyGlnAlaValIleGlnAlaAspGlyValValArgThrArg 129
    |||||
QY 376 CGAAAGTTGAACCTACTTGTGTGACGTACTTGTGTGGGAGGAGGATGTTCCATCG 435
    |||||
DB 130 ArgLysLeuLysProThrHisValGlnArgThrLeuPheGlyGlnGlyAspGlySerAsp 149
    |||||
QY 436 CTAGCGGTTTTCAGACGCTGTGTGGAAAGCTGGGTGCTTATGCTGTTGGAGCACCTT 495
    |||||
DB 150 LeuValIleAspGlnThrSerLeuGlnArgValGlySerLeuGlyCysTrpIleHisLeu 169
    |||||
QY 496 CAACCGCTACAAATAGCCTTGTATGACAAATAGCAAAATGATTCATTTGCGGCTTGG 555
    |||||
DB 170 GlnProLeuThrLysTyrAlaMetCysSerGlnHisGlnGlnIleHisIleAlaIleTrp 189
    |||||
QY 556 CCGACCTTATGACCTTATCTCTAATGCGGGAAGCCCTGGGGCTGATGTCATGTAAGCG 615
    |||||
DB 190 ProSerPheSerIlePheProGlnAlaValTyrAlaLeuGlnGlyProGlnValAsnThrAla 209
    |||||
QY 616 GCCTTCGCAATCTATGCGCTTGAAGGGCAATGCTTCGTAACGCTGCTGCGCTCGTT 675
    |||||
DB 210 AlaSerGlnGlnLysTyrAlaValGlnGlnGlnIleThrTyrValLeuAlaProCysAlaValIle 229
    |||||
QY 676 TCACATCATGATGATATGCTTGTGACAGATGAGCAAAAGCATGCGTGTGCTCGGCT 735
    |||||
DB 230 GlyAspAlaIleGlyTrpIleAlaPheAlaAspThrGlnGlnLysArgGlnLeuIleHisLys 249
    |||||
QY 736 GGTGTGACACTCACTCATGATGAGGCGCTGATGTGTGATGCTGCTGCGCTCTTGGC 795
    |||||
DB 250 GlyGlyGlyTyrAlaArgGlyLeuTyrGlyProAspGlyArgSerLeuAlaGlnProLeuAla 269
    |||||
```

```

QY 796 GAAATGAGAGGGATATTCTCTACGCAACCTTGATCTCGAGTACGATCTGTCTAA 855
    |||||
DB 270 ProAsnAspGlnGlyIleLeuTyrAlaAspIleAspLeuSerAlaIleLeuAlaIleLys 289
    |||||
QY 856 ATGCGCGCAGACCCCTGTGCTCATTTATTCCTGCCGACATCTCGCTGCTTAATGAT 915
    |||||
DB 290 AsnProAlaAspProValGlyHisTyrSerArgProAspValLeuArgLeuGlyPheAsn 309
    |||||
QY 916 CGCAGCCCTTAATTAACCGGTAGTGAATGAAGGT 951
    |||||
DB 310 LysAlaProGln--ProLysValAsnIleLeuGly 320
    |||||
```

## RESULT 2

```

US-09-806-876A-2
; Sequence 2, Application US/09806876A
; Patent No. 6869783
; GENERAL INFORMATION:
; APPLICANT: Kess-Loeschke, Marion
; APPLICANT: Friedlich, Thomas
; APPLICANT: Hauser, Bernhard
; TITLE OF INVENTION: A process for preparing chiral carboxylic acids
; TITLE OF INVENTION: from nitriles using a nitrilase or microorganisms
; TITLE OF INVENTION: which comprise a gene for the nitrilase
; FILE REFERENCE: 49462
; CURRENT FILING DATE: 2001-05-21
; PRIOR APPLICATION NUMBER: Germany/19848129.2
; PRIOR FILING DATE: 1998-10-19
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: WordPerfect version 6.1
; SEQ ID NO 2
; LENGTH: 356
; TYPE: PRT
; ORGANISM: Alcaligenes faecalis
US-09-806-876A-2
```

## Alignment Scores:

Pred. No.:	4,876-82	Length:	356
Score:	822.00	Matches:	160
Percent Similarity:	65.1%	Conservative:	56
Best Local Similarity:	48.2%	Mismatches:	114
Query Match:	45.0%	Indels:	2
DB:	2	Gaps:	1

US-09-751-299-3 (1-1014) x US-09-806-876A-2 (1-356)

```

QY 4 AAAGAGTATGACAGTCGCTGCTGCAAGCCGCCGATCTACATGATTTGAGAGCG 63
    |||||
DB 4 ArgLysIleValAlaGlnAlaAlaValGlnAlaIleAspProAsnTyrAspLeuAlaThr 23
    |||||
QY 64 ACGGTGACAAACCATTTGATGATGAGAGACAGACGATTAATATGCTGCTGATC 123
    |||||
DB 24 GlyValAspLysThrIleGlnLeuAlaArgGlnAlaArgAspGlnGlyCysAspLeuIle 43
    |||||
QY 124 GCCTTCGGAACCTGATTCAGGCTTACCCAGGATGCTTGTGCTGATGATGACGCA 183
    |||||
DB 44 ValPheGlyGlnThrLeuPheProGlnTyrProPheHisValIlePheGlyAlaProAla 63
    |||||
QY 184 TGGGCATGCAATTTGTAACGCCAATACCATGAGAACTATGAGTTGATGATGCTTCAA 243
    |||||
DB 64 TrpSerLeuLysTyrSerAlaArgTyrAlaAsnSerLeuSerLeuAspSerAlaGln 83
    |||||
QY 244 GCTAAGCGCATTTACATGACGCGCAAGCGGTTGGGAATCATGTCACCTTGGGATGAGT 303
    |||||
DB 84 PheGlnArgIleAlaGlnAlaAlaArgThrLeuGlyIlePheIleAlaLeuGlyTyrSer 103
    |||||
QY 304 GAACGGGTGCGTGGACCCCTTACATGATGATGATGATGATGATGATGATGATGATGAT 363
    |||||
DB 104 GlnArgSerGlyGlySerLeuTyrLeuGlnGlnCysLeuIleAspAspLysGlyGlnMet 123
    |||||
QY 364 ATTGGGCGCCGCGGCAAAATGTAACCTTATGTTGTAAGCTTGTGCTGCGGAAAGG 423
    |||||
DB 124 LeuTrpSerArgArgLysLeuLysProThrHisValGlnArgThrValPheGlyGlnGly 143
    |||||
```



```
QY 424 GATGGTCATCGCGGTTTTCAGACGCTGTGGAGGCTGGAGCTTATGCTGT 483
Db 144 TTTAAlaAgArLeuIleValSerAphrThrgluArgValIGlValAlaLeuGlySer 163
QY 484 TGGAGCACTTTCACCGCTTAACAATAACGCTTTGTATGCAACAAATGAAGATTCAT 543
Db 164 TTTGlnHleuSerProleuSerlyThrAlaLeuGlySerGlnHleuAlaLeuHle 183
QY 544 TGTGGCGCTGGCCAGCTTTAGCTTTATCTTAATGCGGCAAGCCCTGGGCGCTGAT 603
Db 184 TleAlaAlaThrProSerPheSerleuGlySerGlnAlaHleAlaLeuSerAlaGly 203
QY 604 GTCAATGATAGCGGCTTCGATCTTATGCGGTTGAAGGCAATGCTTGTACTAGCG 663
Db 204 ValAlaMetAlaAlaSerGlnIleTySerValIGlGlyGlnCyPheThrIleAlaAla 223
QY 664 TGTGGCGCTGTTTCAACATCCATGATCGATATGCTTTGTACAGATGAGAAAGCATCG 723
Db 224 SerSerValIleThrGlnGlnThrleuAserMetleuGluValIGlGlnHleAserPro 243
QY 724 TTTGCTTTGCGCTGGTGGAGACATCGATCATAGGCGCTGATGTGGTGAATTCGTC 783
Db 244 LeuLeuGlyValIGlGlySerSerMetIlePheAlaProAspGlyArgThrleuAla 263
QY 784 GCGCGCTTGGCGGAATGAAGAGGTTATCTCTACGGAACCTTGATCTCGAGTACGC 843
Db 264 ProTyLeuProHleAserAlaGlnIleuLeuIleAlaAserLeuAserGlnIle 283
QY 844 ATCTTGTCTAAATGCGCGCAGACCTCTGCTGCTATTTCCCTCCGACATTCCTGCG 903
Db 284 AlaPheAlaGlyAlaIleAserAserProValIGlGlnHleTySerlyProGlnAlaThrArg 303
QY 904 TTGCTAATAGATCGACGCCCTTAATTCACCGTAGTTGAATGAAGGTGATCTTCCT 963
Db 304 LeuValleuAserPheGlnHleAserAserProMetThrArgValHleSer-----LysSer 321
QY 964 TACGCTTGGGTGAAGCGTGCAGACGCGGTGCGCA 999
Db 322 ValThrArgGluGluAlaProGlnGlnAlaGln 333

RESULT 3
US-09-823-373-5
; Sequence 5, Application US/09823373
; Patent No. 6870038
; GENERAL INFORMATION:
; APPLICANT: Chauhan, Sarita
; APPLICANT: Dicosimo, Robert
; APPLICANT: Payne, Mark
; APPLICANT: Gavigan, John
; TITLE OF INVENTION: Isolation and Expression of a Gene for Nitric Oxide Synthase
; FILE REFERENCE: BC-1032 US NA
; CURRENT APPLICATION NUMBER: US/09/823,373
; PRIOR APPLICATION NUMBER: 60/193,707
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: Microsoft Office 97
; SEQ ID NO: 5
; LENGTH: 369
; TYPE: PRT
; ORGANISM: Acidovorax Facilis
US-09-823-373-5

Alignment Scores:
Pred. No.: 6,128-77 Length: 369
Score: 776.00 Matches: 150
Percent Similarity: 67.04 Conservative: 61
Best Local Similarity: 47.64 Mismatches: 100
Query Match: 42.54 Indels: 4
DB: 2 Gaps: 3
```

```
US-09-751-299-3 (1-1014) x US-09-823-373-5 (1-369)
QY 22 GCTGCGTCAAGCCGCCGATCATATGATTTGGAGCGGACGATGACAAACATT 81
Db 11 AlaThrValGlnAlaGluProValIleProValIleAserProAlaThrIleAserIle 30
QY 82 GAGTTGATGAAGAGGACGATATATGCTGCTGATGCTGCTTCCGGAACCTTG 141
Db 31 GlyIleIleGlnGluAlaAlaGlnLeuGlyAlaSerleuIleAlaPheProGluValPhe 50
QY 142 ATTCAGGCTTACCCGATGCTTTGCTTGTGCTGATCACCAGCATGGCAATTTGTA 201
Db 51 IleProGlyTyProTyThrPheAlaThrleuGluValAserValIleTySerleuSerPheThr 70
QY 202 CGCCAAATACATGAAATCTATGAGTGGATGAGCCCTCAAGCTTAAGCGATTCAGAT 261
Db 71 SerArgTyHleGlnAlaAserleuGlnleuGlyAserAserGlyArgGluGlnleu 90
QY 262 GCACCGAAGCGCTTGGGAATCATGTCACCCCTGGGAGATGATGAACGGCTCGGACCC 321
Db 91 AlaAlaArgAserValIleAlaLeuAlaMetGlyTySerGluArgGluAlaGlySer 110
QY 322 CTTCATACATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 381
Db 111 ArgTyLeuSerGlnValPheIleAserProValIGlGlnIleValAlaAserAserGly 130
QY 382 TTGAACCTTATCTTTGTTGAACGTTATCTTTGCGGCAAGGAGTTCATCGCTAGCG 441
Db 131 LeuAserProThrHleValIGlGlnArgThrIleTyGlnIleGlnAlaGlnIleAserPheleu 150
QY 442 GTTTGAGACGCTGTGTTGAAGCGTGGGCTTATGCTGTTGAGGACCTTCACCG 501
Db 151 ThrHleAserPheAlaPheGlyArgValIGlGlyLeuAserCyThrGlnHlePheGlnPro 170
QY 502 CTACAAATATGACCTTTGATGACCAAAATGAAGAGATTCATTTGGGCTGGCGGAC 561
Db 171 LeuSerlyPheMetMetTySerleuGlnIleValHleValAlaAserThrProAla 190
QY 562 TTTAGC--CTTATCTTAAATGCGCGCAAGCCCTGGGCGCTGATGTCATGACGCGC 618
Db 191 MetSerProleuGlnProAspValPheGln--LeuSerIleGlnAlaAserAlaThrVal 209
QY 619 TCTGATCTATGCGCTTGAAGGCAATGCTTGTATACGCTGCTGCTGCTGCTTCA 678
Db 210 ThrArgSerTyAlaIleGlnIleGlnIleThrPheValleuCySerThrGlnValIleGly 229
QY 679 CAATCCATGATCGATATGCTTTGTACAGATGACGAAAGCATGGCTTGGCTGGCTGT 738
Db 230 ProSerAlaIleGlnThrPheCyAserAserProIleGlnAlaLeuLeuProGlnGly 249
QY 739 GTTGACACTCAGCTATCATAGGCGCTGATGCTGCTGCTGCTGCTGCTGCTGCGGAA 798
Db 250 CysGlyTyThrAlaArgIleTyGlyProAspGlySerGlnleuAlaGlyProleuAlaGln 269
QY 799 AATGAAGAGGATTTCTTACGCAACCTTGATCTGAGTACGATCTTGTCTAAATG 858
Db 270 AspAlaGlnIleGlyIleleuTyAlaGlnIleAserleuGlnIleleuAlaAlaAla 289
QY 859 GCGGACAGCCCTGCTGCTTATTCCTGCGGACATTCGCTGCTGCTGCTGCTGCTGCTG 918
Db 290 GlyAlaAserProValIGlGlnIleTySerArgProAserAlaLeuSerValGlnPheAserPro 309
QY 919 AGCCCTTAATTCAGGCTA-----GTTGAATTCGAAGTATCTT 957
Db 310 ArgAserHleThrProValHleAserGlnIleGlyIleAserProIleArgleu 324

RESULT 4
US-09-823-373-14
; Sequence 14, Application US/09823373
; Patent No. 6870038
; GENERAL INFORMATION:
; APPLICANT: Chauhan, Sarita
```



## Alignment Scores:

Pred. No.: 1,14e-61 Length: 354  
 Score: 638.00 Matches: 134  
 Percent Similarity: 60.3% Conservative: 62  
 Best Local Similarity: 41.2% Mismatches: 107  
 Query Match: 34.9% Indels: 22  
 DB: 1 Gaps: 5

US-09-751-299-3 (1-1014) x US-08-447-702-5 (1-354)

```

QY      13 ATCAAGTTCGCTGCGTGCAGAACCGCCCGATCTACATGATTTGGAGCGACGGTGAC 72
       7 VallysValAlaAlaValAlaGlnAlaAlaProValPheMetAsnLeuGlnAlaThrValAsp 26
Db      73 AAAACATTGAGTTGATGGAAGACAGACAGTAATATGCTGCTGATGCGCTTTCCG 132
       27 LysThrCysLysLeuIleAlaGlnAlaAlaSerMetGlyAlaLysValIleGlyPhePro 46
QY      133 GAAACTTGAGTCCAGGCTACCCATGCTTTCTTTGGCTTGACTCA----- 177
       47 GluAlaPheIleLeuProGlyLysProGlyTrpIleTrpIleThrSerAsnMetAspPheThrGly 66
Db      178 CCAGCATGGGCAATGCAATTTGTACGCCAATACATGAGAACTCATTTGAGTTGATGGAC 237
       67 MetMetTrpAlaValLeuPhe-----LysAsnAlaIleGlnIleProSer 81
QY      238 CCTCAAGCTAAGCCGATTTGATGACAGCGCAAGCGGTTGGAGATGATGCTCCTGGAG 297
       82 LysGlnValGlnGlnIleSerAspAlaAlaLysLysAsnGlyValLysValCysValSer 101
QY      298 ATGAGTGAACGGGCGGTGGGCGGCTTTACATCAGTCAGTGGTTCATAGGCGATATAGT 357
       102 ValSerGlnLysAspAsnAlaSerLeuLysLeuTrpGlnLeuTrpPheAspProAsnGly 121
Db      358 GACACCATTTGGGCGCGCGGAAAGTTGAACCTACTTTTGTGAACGTATTGTTCCGAC 417
       122 AsnLeuIleGlyLysHisArgLysPheLysProThrSerSerGlnAlaAlaValIleProGly 141
QY      418 GAAAGGAGATGGTTTCATCCGCTAGCGGTTTTCGAGACGTCGTGTGGAAGCGGTGGGCTTA 477
       142 AspGlyAspGlySerMetAlaProValPheLysThrGlnLysGlyAsnLeuIleGlyLeu 161
Db      478 TGCCTGTGGGAGCACCTTCAACCGCTTAACAAATACGCTTGTATGACAAATGAATGAG 537
       162 GlnCysTrpGlnHisAlaLeuProLeuAsnIleAlaAlaMetGlySerLeuAsnGln 181
QY      538 ATTCAATTGCGCGGCTTGGCGGAGCTTTAGCCTTTATCTCAATGCGGCAAGCCCTGGAG 597
       182 ValHisValAlaSerTrpProAlaPhe-----ValProLysGlyAlaValSerSerArg 199
Db      598 CCTGATGTCATGTAGCGGCC-----TTCGAATCTAT 630
       200 ValSerSerSerValCysAlaSerThrAsnAlaMetHisGlnIleIleSerGlnPheLys 219
QY      631 GCCGTTGAAGGCGCATGCTTCTACTAGCGTCGTGGCGCTTCGTTCAACATCCATGATC 690
       220 AlaIleSerAsnGlnValLysValIleMetSerThrAsnLeuValGlyGlnAspMetIle 239
Db      691 GATATGCTTTGTACAGATGACGAAAGACATGCGTCTTCTGCGTGGGAGTGGACACCTCA 750
       240 AspMetIleGlyLysAspGlnPheSerLysAsnPheLeuProLeuGlySerGlyAsnThr 259
QY      751 CGTATCATAGGGCGCTGATGGTGTGATGCTGGCGGCTTTGCCGAAATGAAGAGGT 810
       260 AlaIleIleSer--AsnThrGlyGlnIleLeuAlaSerIleProGlnAspAlaGlnGly 278
Db      811 ATTCTCTACGCAACCTGATTCCTGAGAGACATCCCTTGCTAAATAGCGGCAACCT 870
       279 IleAlaValAlaGlnIleAspLeuAsnGlnIleLysGlyLysTrpLeuLeuAspPro 298
QY      871 GCTGCTATTATTCGCGCGTCCGACATTAATGCTGCTGATGATGCGGCGCCTTAATTA 930
       299 AlaGlyHisLysSerThrProGlyPheLeuSerLeuTrpPheAspGlnSerGlnHisVal 318

```

QY 931 CCGGATGTTGAATT 945  
 Db 319 ProValLysLysIle 323

RESULT 6  
 US-08-465-615-5

Sequence 5, Application US/08465615  
 Patent No. 5635391

GENERAL INFORMATION:

APPLICANT: PETRE, Dominique  
 APPLICANT: CERBELAUD, Edith  
 APPLICANT: LEVY-SCHUL, Sophie  
 APPLICANT: CROUZET, Joel  
 TITLE OF INVENTION: POLYPEPTIDES POSSESSING A NITRILASE  
 TITLE OF INVENTION: ACTIVITY, DNA SEQUENCE CODING FOR SAID POLYPEPTIDES,  
 TITLE OF INVENTION: EXPRESSION CASSETTES AND HOST MICROORGANISMS ENABLING THEM  
 TITLE OF INVENTION: TO BE OBTAINED, AND METHOD OF CONVERTING NITRILES TO  
 NUMBER OF SEQUENCES: 5  
 CORRESPONDENCE ADDRESS:  
 ADDRESSEE: Burns, Doane, Swecker & Mathis  
 STREET: P.O. Box 1404  
 CITY: Alexandria  
 STATE: Virginia  
 COUNTRY: United States  
 ZIP: 22313-1404

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk  
 COMPUTER: IBM PC compatible  
 OPERATING SYSTEM: PC-DOS/MS-DOS  
 SOFTWARE: Patent in Release #1.0, Version #1.25  
 CURRENT APPLICATION DATA:  
 APPLICATION NUMBER: US/08/465,615

FILING DATE:

CLASSIFICATION: 435  
 PRIOR APPLICATION DATA:  
 APPLICATION NUMBER: US 08/194,588

FILING DATE: 10-FEB-1994

APPLICATION NUMBER: FR 9209882

FILING DATE: 10-AUG-1992

ATTORNEY/AGENT INFORMATION:

NAME: Crane-Feuzy, Sharon E

REGISTRATION NUMBER: 36,113

REFERENCE/DOCKET NUMBER: 003025-015

TELECOMMUNICATION INFORMATION:

TELEPHONE: (703) 836-6620

TELEFAX: (703) 836-2021

INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:

LENGTH: 354 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

US-08-465-615-5

## Alignment Scores:

Pred. No.: 1,14e-61 Length: 354  
 Score: 638.00 Matches: 134  
 Percent Similarity: 60.3% Conservative: 62  
 Best Local Similarity: 41.2% Mismatches: 107  
 Query Match: 34.9% Indels: 22  
 DB: 1 Gaps: 5

US-09-751-299-3 (1-1014) x US-08-465-615-5 (1-354)

```

QY      13 ATCAAGTTCGCTGCGTGCAGAACCGCCCGATCTACATGATTTGGAGCGACGGTGAC 72
       7 VallysValAlaAlaValAlaGlnAlaAlaProValPheMetAsnLeuGlnAlaThrValAsp 26
Db      73 AAAACATTGAGTTGATGGAAGACAGACAGTAATATGCTGCTGATGCGCTTTCCG 132
       27 LysThrCysLysLeuIleAlaGlnAlaAlaSerMetGlyAlaLysValIleGlyPhePro 46

```





ATTORNEY/AGENT INFORMATION:  
NAME: Player Esq., William E.  
REGISTRATION NUMBER: 31,409  
REFERENCE/DOCKET NUMBER: P-500-23486  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-887-0400  
TELEFAX: 202-835-0605  
TELEX: 440706  
INFORMATION FOR SEQ ID NO: 5:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 311 amino acids  
TYPE: AMINO ACID  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-07-917-111-5

Alignment Scores:  
Pred. No.: 1,866-06 Length: 311  
Score: 139.00 Matches: 80  
Percent Similarity: 38.3% Conservative: 49  
Best Local Similarity: 23.7% Mismatches: 146  
Query Match: 7.6% Indels: 62  
DB: Gaps: 15

US-09-751-299-3 (1-1014) x US-07-917-111-5 (1-311)

QY 13 ATCAAGTCGCTGCTGCAAGCCGCCGATCTACATGAT-----TTGAGCGCAGC 66  
DB 4 Valaenlaalaalalaglmecglprrlleserargluthrarglyasprthr 23  
QY 67 GTGACAAACCATGATGATGGAAGAAGACAGACCAATTAATGTCGTCTGATCGCC 126  
DB 24 Valaaglaagllellaaleumetckrglualalysalaaglyseraspleuvalal 43  
QY 127 TTTCCGAAACTGGATTCAGAGTCCAGGTTCCTTGGCTTCACTCACAGCATGG 186  
DB 44 Phehrhrglueuallaaleuthrphr-----Phehrargtrp 56  
QY 187 GCAATGCAATTGTACGCCATACATGAGAACTGATGGATTGGAT----- 234  
DB 57 Valile-----gluaprgluagluleasprpethrgrlu 69  
QY 235 -----GGCCTCAAGCTAAGCGCATTTGATGACGACCAAGCGTTGGGAATC 282  
DB 70 Lysglumetprrgluprrgluthrghlnprrleuphrpabrgluallblyhrgleuulu 89  
QY 283 ATGTCACCCCTGGGAGATGAGTAA-----CGGTCGGTGGC-----ACCTTTACATC 330  
DB 90 Glypethrgrleuaglytrgluagluagluagluaglyarglyasprghaspr 109  
QY 331 AGTCAGTGGTTCAATAGCGGATTAATGTCACACCATTTGGGGCGGGAAGTTG----- 384  
DB 110 ThrSerlleleuValaAsprgserdlyargllevalglylgrarglyasprhlsleu 129  
QY 385 -----AAACCTACTTTTGGTGAACGTAATGCGGCTTTGTTTC 414  
DB 130 Prrgluylslygluprrgluprrgluprrgluylslygluylsleuaglylgrarglyr 149  
QY 415 GCGGAGGAGATGTTTCATCGCTAGCGGTTTTCGAGACGCTGTGTGAAAGCTGGTGGC 474  
DB 150 GluprrgluAsp---leuaglyrhegllyaltrparglarpheasprgllyalmetglumet 168  
QY 475 TTATGCTGTGGGAGACACTTCAACCGGTAACAAATACGCTTTGATGACCAAAATGAA 534  
DB 169 CyslleCysaenAsprargtrgtrprrgluuthrlyrargvalmetglyleuindlyval 188  
QY 535 GAGATTCATTTGCGCGCTTGG-----CCGACCTTTAGGCTTTATCTTAATGCGCGAA 588  
DB 189 GlumetValmetleuaglytrarghlnprrgluylsleuaglylgrarglyasprlleaspr 208  
QY 589 GGCCTGGGGGCTGATGTCATGTAAGGCGCTTCGATATGATCGCGTGAAGGCGAAATGC 648  
DB 209 SerleuthrghlnprrhlsasprhlsleuSermetglualaglyalarglylinsprthr 228

QY 649 TTGTACTAGCGHGTGTGCGCTGTTTGCATTCATGATGATATGTTTGAAGAT 708  
DB 229 trpvalleelythrlnalys-----Cysglythr 238  
QY 709 GACAAAGCATCGTTGCTTCTGCTGTGTGACACTGACATGATATGAGGCGCTGAT 768  
DB 239 GluGluylserlysmetVal-----GlylnserVallevalalapro--- 253  
QY 769 GGTGTGACTTGTGCGGCGCTTTCGCGAAATGAAGGATTTCTTACGCA----- 822  
DB 254 SerdlygluilevalalmetalaCysthrlllegluasprlluilethrlnalargCys 273  
QY 823 AACCTGATCCTGAGACGATCCTTGTCTAATGATGCGAGCCCTTAATTCAG 882  
DB 274 Aspleuasprmetgllylsarg---Tyrarggluthrllheasprhealarghls--- 291  
QY 883 TCCCGTCCGACATTAATCTGCTGCTAATGATGCGAGCCCTTAATTCAG 933  
DB 292 Arggluprraspralatrargleuilevalgluarglyasgllyalvalpro 308

RESULT 10  
US-07-917-111-6  
Sequence 6, Application US/07917111  
Patent No. 5565344  
GENERAL INFORMATION:  
APPLICANT: Namba, Hirokazu  
APPLICANT: Yamada, Yukio  
APPLICANT: Takano, Masayuki  
APPLICANT: Ikenaka, Yasuhiro  
APPLICANT: Takahashi, Satomi  
APPLICANT: Yajima, Kazuyoshi  
TITLE OF INVENTION: PROCESS FOR PRODUCTION OF D-ALPHA-AMINO  
NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESS: Wegner, Cantor, Mueller & Player  
STREET: 1233 20th Street, N.W., Suite 300  
CITY: Washington  
STATE: D.C. 20036-8218  
ZIP: 20036-8218  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/07/917,111  
FILING DATE: 19920807  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 400848/1990  
FILING DATE: 07-DEC-1990  
APPLICATION NUMBER: JP 407922/1990  
FILING DATE: 27-DEC-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 078840/1991  
FILING DATE: 11-APR-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/JP91/01696  
FILING DATE: 06-DEC-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Player Esq., William E.  
REGISTRATION NUMBER: 31,409  
REFERENCE/DOCKET NUMBER: P-500-23486  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-887-0400  
TELEFAX: 202-835-0605  
TELEX: 440706  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 311 amino acids



Alignment Scores:

Pred. No.:	1.86e-06	Length:	311
Score:	139.00	Matches:	80
Percent Similarity:	38.3%	Conservative:	49
Best Local Similarity:	23.7%	Mismatches:	146
Query Match:	7.6%	Indels:	62
DB:	1	Gaps:	15

US-09-751-299-3 (1-1014) x US-08-479-638-5 (1-311)

```
QY      13 ATCAAGTCGCGCTGCTGCAAGCCGCGCCGATCTACATGAT-----TTGAGGCGACG 66
      4 ValasnaalaalaalaglmecglYproilesersergluthrArglyasprhr 23
QY      67 GTGACAAAACATTGATGATGAGAAGACGACGTAATATGCTCGTGTATGCC 126
      24 Valargargleuilealeuemetarglualalyalarglyseraspleuvala 43
QY      127 TTTCCGGAACCTTGATTCAGGCTACAGGCTACGATTTCTTGGCTTGACCTACGACATGG 186
      44 PhehrngluleuilaaleuthrThrPhe-----PheProhArgTtr 56
QY      187 GCAATGCAATTGTGACGCCAATACCATAGAACTCATTTGAGTTGAT----- 234
      57 Valile-----gluaspglualagluleuaspserPheTyrGlu 69
QY      235 -----GGCCCTCAAGCTAAGCGCATTTCAATGACAGCAAGCGGTTGGAGATC 282
      70 LysglumetProglYprogluhrnglInProleuPheasprglualalyasrgleuclule 89
QY      283 ATGTCACCTCGGGGATGATGAA-----CGGGTCGGTGGC-----ACCTTTACATC 330
      90 GlyPheTyrleuaglYtyralagluleuilaagluluglYarglyasrglyasrPheasn 109
QY      331 AGTCAGTGTTCATAGCGCATATGTGTACACACCATGGGGCGCGCAAGTTG----- 384
      110 ThrSerileuValasprargserlYarglilevalglYsYtyrarglyvalnlsleu 129
QY      385 -----AAACCTACTTTTGTGTGAACGTAATCTTGTTC 414
      130 ProglYnlsYsYgluProglInProglYarglyshlsYnlsYleuclYlYarglyrPhe 149
QY      415 GCGCAAGGGGATGTTTCATCGCTACGCGTTTCGAGACGTCGTTTGAAGCGCGTGGC 474
      150 GluProglYaspr---leuGlYrPheglYvallYrpararglaphasprglYvalmet 168
QY      475 TTATCTCTGTGGGACACCTTCACCGGCTAACAAATACGCTTTGTATGCAAAATGAA 534
      169 CyslleCysasnaspargargTtrProgluThrTyrargvalmetGlyleuclnlyval 188
QY      535 GAGATTCATTGTGCGCGCTTG-----CCGAGCTTTAGCCTTTATCCTAATCGCGCAAA 588
      189 GlumetValmetleuclYtyrAsnhrProtyrAsnhrsthrnglYnlsYasprlsleasr 208
QY      589 GCCCTGGGCGCTGATGTATGATGTAGCGGCTCGAATCTATGCCGTTGAAGGCGAATGC 648
      209 SerleuThrnglInPhehlsAsnhsleuSermetGlnalaglYalatyGlnasrThr 228
QY      649 TTGCTACTAGGCGTGTGCGCGCTGTTTCAACATCCATGATGATATGCTTTGTACAGAT 708
      229 TtrValilleglYthAlaLys-----CysglYthr 238
QY      709 GACGAAAACATGCGCTGCTCTTGCTGCTGCTGCTGACATCAGCTATCATAGGCGCTGAT 768
      239 GluGlulYserlyMetVal-----GlylnserValillevalAlaPro--- 253
QY      769 GGTGTGACTTGCTGTGCGCGCTCTTCCGAAAATGAAGAGGATTTCTTACGCA----- 822
      254 SerglYgluilevalAlaMetAlaCysThrillegluaspgluilelthrAlaargCys 273
QY      823 AACCTGATCTCGAGTACGATCTCTGCTAATAAGCGGACGCGCTGCTGATCATAT 882
      274 AspLeuaspmetGlylYarg---TyrarggluThrillePheaspPheAlaarghls--- 291
```

QY 883 TCCCGTCCGACATTACTCGCTGCTAATAGATCGACGCCCTTAATTACCG 933  
Db 292 ArggluProaspAlaTyrArgyleuileValgluarglyglYalavalPro 308

RESULT 12

US-08-479-638-6

Sequence 6, Application US/08479638

Patent No. 5695968

GENERAL INFORMATION:

APPLICANT: Nanda, Hirokazu

APPLICANT: Yamada, Yukio

APPLICANT: Takano, Masayuki

APPLICANT: Ikenaka, Yasuhiro

APPLICANT: Takahashi, Satomi

APPLICANT: Yajima, Kazuyoshi

TITLE OF INVENTION: PROCESS FOR PRODUCTION OF D-ALPHA-AMINO

TITLE OF INVENTION: ACIDS

NUMBER OF SEQUENCES: 6

CORRESPONDENCE ADDRESS:

ADDRESSER: Wegner, Cantor, Mueller & Player

STREET: 1233 20th Street, N.W., Suite 300

CITY: Washington

STATE: D.C.

ZIP: 20036-8218

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/479,638

FILING DATE: 07-JUN-1995

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 07/917,111

FILING DATE: 07-AUG-1992

APPLICATION NUMBER: JP 400848/1990

FILING DATE: 07-DEC-1990

PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP 407922/1990

FILING DATE: 27-DEC-1990

PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP 078840/1991

FILING DATE: 11-APR-1991

PRIOR APPLICATION DATA:

APPLICATION NUMBER: PCT/JP91/01696

FILING DATE: 06-DEC-1991

ATTORNEY/AGENT INFORMATION:

NAME: Player Esq., William E.

REGISTRATION NUMBER: 31,409

REFERENCE/DOCKET NUMBER: P-500-23486

TELECOMMUNICATION INFORMATION:

TELEPHONE: 202-887-0400

TELEFAX: 202-835-0605

TELEX: 440706

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 311 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: protein

ORIGINAL SOURCE:

ORGANISM: Pseudomonas

STRAIN: KMK 003A (FERM BP-3181)

US-08-479-638-6

Alignment Scores:

Pred. No.:	1.86e-06	Length:	311
Score:	139.00 <td>Matches:</td> <td>80</td>	Matches:	80
Percent Similarity:	38.3% <td>Conservative:</td> <td>49</td>	Conservative:	49
Best Local Similarity:	23.7% <td>Mismatches:</td> <td>146</td>	Mismatches:	146





```

; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 672-5300
; TELEFAX: (202) 672-5399
; TELEX: 904136
; INFORMATION FOR SEQ ID NO: 70:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 311 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-294-871A-70

Alignment Scores:
Pred. No.: 1.86e-06 Length: 311
Score: 139.00 Matches: 80
Percent Similarity: 38.3% Conservative: 49
Best Local Similarity: 23.7% Mismatches: 146
Query Match: 7.6% Indels: 62
DB: Gaps: 15

US-09-751-299-3 (1-1014) x US-08-294-871A-70 (1-311)

QY 13 ATCAAGTCGCGCTGCGTCAGACCGCCGATCTACATGAT-----TTGAGGCGACG 66
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 4 ValasnlaalaalaalaglmecglyProilesergsergluthrArglyasprthr 23

QY 67 GTGACAAACACCATTTGATGTGAGAGACAGACGTAATTAATGCTGCTGATGCGC 126
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 24 ValArgaglyleuilealaaleumecArglyualalysalaargglyserAspIeuVal 43

QY 127 TTTCCGGAACCTTGATTCAGGCTACCCATGTTCTTTGGCTTGACTCAGCATGG 186
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 44 PheThrGluLeuAlaLeuThrPhe-----PheProArgIrp 56

QY 187 GCATGCAATTTGACGCCATACATAGAACATCATGAGTTGGAT----- 234
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 57 ValIle-----GluAspGluAlaGluLeuAspSerPheTyrglu 69

QY 235 -----GGCCCTCAAGCTAAGCGCATTTGCATGACGACCAACGGTTGGGAATC 282
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 70 LysGluMetProglyProgluThrGlnProleuPheAspGlyualalysArgleuIle 89

QY 283 ATGCTACCTGGGGATGATGAA-----CGGGTCGGTGGC-----ACCTTTATAC 330
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 90 GlyPheTyrlleuglyTyralagluIeuAlaIugluIglyArglyAspArgPheAsn 109

QY 331 AGTCAGTGTTCATACAGCGGATATATGGACACCATTTGGGGCCCGCGAAAGTG----- 384
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 110 ThierileleuValAspArgserIyArgIleValGlyLysTyArgLysValIleIeu 129

QY 385 -----AACTACTTTTGTGAACGTAATTGTTTC 414
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 130 ProgluHisLysGluProgluProgluArgLysHisGlnHisleuGluLysArgTyrrhe 149

QY 415 GCGCAAGGAGATGTTTCATGCGCTAGCGGTTTTCGAGAGCTGCTGTGAAGCTGGGTCG 474
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 150 GluProGlyAsp---LeuGlyPheGlyValAlIrraGAlaPheAspGlyIuMetIyMet 168

QY 475 TTATGCTGTGGAGACCTTCAACCGCTTAACAATAATAGCTTTATGACCAATAATGAA 534
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 169 CysIleCysAsnAspArgArgIrrProgluThrTyArgValIuMetGlyIuendIuIyAl 188

QY 535 GAGATTCATTGTGGCGGCTTG-----CCGAGCTTTAGCCTTTATCTTAATGGCGGAAA 588
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 189 GluMetValMetIeuGlyTyrrAsnThrProTyrrAspHisIthrGlyHisIAspAspIleAsp 208

QY 589 GCGCTGGGGCGCTGATGATAGTAGAGCGCTCTGAATCTATGCGCTGGAAGGCGAAATGC 648
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 209 SerIeuThrGlnPheHisAsnHisIleuSerMetGlnAlaGlyAlaTyrrGlnAsnSerThr 228

QY 649 TTCGTAAGGCTGCTGCGCTGCTTTCACAAATCATGATGATATGCTTTGTAACAGAT 708
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 229 TrpValIleGlyThrAlaLys-----CyselTyrr 238

```

```

QY 709 GAGCAAAAGCATGCGTTCCTTCTGCGTGTGTGACACTACGATATCATAGGCGCTGAT 768
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 239 GluIuGlySerIySwetVal-----GlyIuSerValIleValAlaPro--- 253

QY 769 GGTGTGACTTGTGTGCGCGCTCTTGCGGAAAATGAGAGGGTATTTCTACGCA----- 822
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 254 SerGlyGluIleValAlaMetAlaCyeThrIleGluAspGluIleIleThrAlaArgCys 273

QY 823 AACCTTATCTCGAGTAGACGATCCTTGCTAAATGGCGGACGACCTGCTGCTGATTA 882
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 274 AspLeuAspMetGlyLysArg---TyrrArgIuThrIlePheAspPheAlaArgHis--- 291

QY 883 TCCCGTCCGACATTACTCGCTTGCTAATAGATGACGACCCCTTAATTACCG 933
   |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db 292 ArgIuProAspAlaTyrrArgIleuIleValGluArgLysGlyAlaValPro 308

RESULT 14
US-08-876-398A-70
; Sequence 70, Application US/08876398A
; Patent No. 6083752
; GENERAL INFORMATION:
; APPLICANT: IKENAKA, Yasuhiro
; APPLICANT: NAKABA, Hirokazu
; APPLICANT: TAKANO, Masayuki
; APPLICANT: YAJIMA, Kazuyoshi
; APPLICANT: YAMADA, Yukio
; APPLICANT: TAKAHASHI, Satomi
; TITLE OF INVENTION: DNA CODING FOR DECARBAMYLASE IMPROVED IN
; TITLE OF INVENTION: THERMOSTABILITY AND USE THEREOF
; NUMBER OF SEQUENCES: 70
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FOLEY & LARDNER
; STREET: 3000 K Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/876,398A
; FILING DATE: 16-JUN-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/211,641
; FILING DATE: 11-APR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: WO PCT/JP93/01101
; FILING DATE: 05-AUG-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 340078/1992
; FILING DATE: 21-DEC-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 212692/1992
; FILING DATE: 10-AUG-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Wegner, Harold C.
; REGISTRATION NUMBER: 25,258
; REFERENCE/DOCKET NUMBER: 74129/130
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 672-5300
; TELEFAX: (202) 672-5399
; INFORMATION FOR SEQ ID NO: 70:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 311 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-876-398A-70

```

[illegible]

```

Db      27  AspleuaspMecGjLyarkY---TyrArgIuThrIlePheAspPheAlaArgHis--- 291
Cy      883  TCCGCTCCCGACATTACTGCTTGTCTAATAGATGCGACCCCTTAATTACCG 933
Db      292  ArgIuPtoAspAlaTyrArgIuLeuValGluHArgIysGlyAlaValPro 308

RESULT 15
US-08-294-871A-46
; Sequence 46, Application US/08294871A
; Patent No. 5824522
; GENERAL INFORMATION:
; APPLICANT: Ikenaka, Yasuhiro
; APPLICANT: Namba, Hirokazu
; APPLICANT: Takano, Masayuki
; APPLICANT: Tajima, Kazuyoshi
; APPLICANT: Yamada, Yukio
; APPLICANT: Takahashi, Satomi
; APPLICANT: Okubo, Kazuma
; APPLICANT: Yamada, Kazuhiko
; APPLICANT: Hiraishi, Yoshiko
; TITLE OF INVENTION: Immobilized Enzyme Preparation and
; TITLE OF INVENTION: Process for Producing D-a-Amino Acid
; NUMBER OF SEQUENCES: 70
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Foley & Lardner
; STREET: 3000 K Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.
; ZIP: 20007-5109
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/294,871A
; FILING DATE: 22-AUG-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/971,758
; FILING DATE: 12-APR-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/917,111
; FILING DATE: 07-AUG-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/211,641
; FILING DATE: 11-APR-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 400848/1990
; FILING DATE: 07-DEC-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP91/01696
; FILING DATE: 06-DEC-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 407922/1990
; FILING DATE: 27-DEC-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 078840/1991
; FILING DATE: 11-APR-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 140051/1991
; FILING DATE: 12-JUN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP92/00739
; FILING DATE: 10-JUN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: JP 212692/1992
; FILING DATE: 10-AUG-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/JP93/01101
; FILING DATE: 05-AUG-1993
; PRIOR APPLICATION DATA:

```



GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - protein search, using frame\_plus\_n2p model

Run on: April 27, 2006, 01:38:11 ; Search time 31.9743 Seconds  
(without alignments)  
2650.121 Million cell updates/sec

Title: US-09-751-299-3

Perfect score: 1826  
Sequence: 1 atgaagaagatcatcaagtc.....cgcactcgagaatttga 1014

Scoring table:

BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 3735138

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Command line parameters:

-MODBL=frame+ n2p model -DEV=xlp  
-O=/abs/ABSSWB.spool/US09751299/runat\_26042006\_090042\_18780/app\_query.fasta\_1  
-DB=Published Applications AA Main -OPMT=fastan -SUFFIX=rapbm -MINMATCH=0.1  
-LOOPCL=0 -LOOPEXT=0 -UNITS=bits -SPART=1 -BND=1 -MATRIX=bloms62  
-TRANS=human40.cdi -LIST=45 -DOCALLIGN=200 -THR\_SCORE=pct -THR\_MAX=100  
-THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=plo -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MAKLEN=200000000 -HOST=abs602p  
-USER=US09751299 @CGN 1 1 405 @runat\_26042006\_090042\_18780 -NCPU=6 -ICPU=3  
-NO MAP -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG -DEV TIMEOUT=120  
-WARN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6 -FGAPEXT=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :

1: /cgn2\_6/ptodata/1/pubpaa/US07\_PUBSCOMB.pep.\*  
2: /cgn2\_6/ptodata/1/pubpaa/US08\_PUBSCOMB.pep.\*  
3: /cgn2\_6/ptodata/1/pubpaa/US09\_PUBSCOMB.pep.\*  
4: /cgn2\_6/ptodata/1/pubpaa/US10A\_PUBSCOMB.pep.\*  
5: /cgn2\_6/ptodata/1/pubpaa/US10B\_PUBSCOMB.pep.\*  
6: /cgn2\_6/ptodata/1/pubpaa/US11\_PUBSCOMB.pep.\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1765	96.7	337	3	US-09-751-299-4
2	1761	96.4	337	4	US-10-146-772-386
3	1761	96.4	337	4	US-10-241-742-386
4	1761	96.4	337	4	US-10-440-523-386
5	1761	96.4	337	4	US-10-440-503-386
6	1761	96.4	337	4	US-10-461-925-386
7	1339	73.3	337	4	US-10-146-772-48
8	1339	73.3	337	4	US-10-146-772-58
9	1339	73.3	337	4	US-10-241-742-48
10	1339	73.3	337	4	US-10-241-742-58
11	1339	73.3	337	4	US-10-440-523-48

12	1339	73.3	337	4	US-10-440-523-58	Sequence 58, App1
13	1339	73.3	337	4	US-10-440-503-48	Sequence 48, App1
14	1339	73.3	337	4	US-10-440-503-58	Sequence 48, App1
15	1339	73.3	337	4	US-10-461-925-48	Sequence 48, App1
16	1339	73.3	337	4	US-10-461-925-58	Sequence 58, App1
17	1139	62.4	333	4	US-10-146-772-292	Sequence 292, App
18	1139	62.4	333	4	US-10-241-742-292	Sequence 292, App
19	1139	62.4	333	4	US-10-440-523-292	Sequence 292, App
20	1139	62.4	333	4	US-10-440-503-292	Sequence 292, App
21	1139	62.4	333	4	US-10-461-925-292	Sequence 292, App
22	1136	62.2	332	4	US-10-146-772-140	Sequence 140, App
23	1136	62.2	332	4	US-10-241-742-140	Sequence 140, App
24	1136	62.2	332	4	US-10-440-523-140	Sequence 140, App
25	1136	62.2	332	4	US-10-440-503-140	Sequence 140, App
26	1136	62.2	332	4	US-10-461-925-140	Sequence 140, App
27	1091	59.7	355	4	US-10-146-772-306	Sequence 306, App
28	1091	59.7	355	4	US-10-241-742-306	Sequence 306, App
29	1091	59.7	355	4	US-10-440-523-306	Sequence 306, App
30	1091	59.7	355	4	US-10-440-503-306	Sequence 306, App
31	1091	59.7	355	4	US-10-461-925-306	Sequence 306, App
32	1081	59.2	345	4	US-10-146-772-334	Sequence 334, App
33	1081	59.2	345	4	US-10-241-742-334	Sequence 334, App
34	1081	59.2	345	4	US-10-440-523-334	Sequence 334, App
35	1081	59.2	345	4	US-10-440-503-334	Sequence 334, App
36	1081	59.2	345	4	US-10-461-925-334	Sequence 334, App
37	1051	57.6	334	4	US-10-146-772-6	Sequence 6, App1
38	1051	57.6	334	4	US-10-241-742-6	Sequence 6, App1
39	1051	57.6	334	4	US-10-440-523-6	Sequence 6, App1
40	1051	57.6	334	4	US-10-440-503-6	Sequence 6, App1
41	1051	57.6	334	4	US-10-461-925-6	Sequence 6, App1
42	1042	57.1	353	4	US-10-146-772-212	Sequence 212, App
43	1042	57.1	353	4	US-10-241-742-212	Sequence 212, App
44	1042	57.1	353	4	US-10-440-523-212	Sequence 212, App
45	1042	57.1	353	4	US-10-440-503-212	Sequence 212, App

#### ALIGNMENTS

RESULT 1  
US-09-751-299-4  
; Sequence 4, Application US/09751299  
; Patent No. US20020012974A1  
; GENERAL INFORMATION:  
; APPLICANT: Madden, Mark  
; APPLICANT: Weiner, David P.  
; TITLE OF INVENTION: METHODS FOR PRODUCING ENANTIOMERICALLY PURE  
; FILE REFERENCE: DIVER1440-2  
; CURRENT APPLICATION NUMBER: US/09/751,299  
; PRIOR FILING DATE: 2000-12-29  
; PRIOR APPLICATION NUMBER: 60/254,414  
; PRIOR FILING DATE: 2000-12-07  
; PRIOR APPLICATION NUMBER: 60/173,609  
; NUMBER OF SEQ ID NOS: 4  
; SOFTWARE: PatentIn Ver. 2.1  
; SEQ ID NO 4  
; LENGTH: 337  
; TYPE: PRT  
; ORGANISM: Unknown Organism  
; FEATURE:  
; OTHER INFORMATION: Description of Unknown Organism: Obtained from an  
US-09-751-299-4  
Alignment Scores:  
Pred. No.: 2.03e-177  
Score: 1765.00  
Percent Similarity: 100.0%  
Best Local Similarity: 100.0%  
Query Match: 96.7%  
DB: 3  
Length: 337  
Matches: 337  
Conservative: 0  
Mismatch: 0  
Indels: 0  
Gaps: 0

US-09-751-299-3 (1-1014) x US-09-751-299-4 (1-337)

QY 1 ATGAAGAAGCTATCAAGGTGCGCTGCTGCAAGCCGCCGATCTACATGATTTGGAG 60  
Db 1 MetylsglnatlellysvalaIaCysvalGlnlaIaIaProileTyrMetAspLeuGlu 20  
QY 61 GCGACGGTGGCAAAACATTGATTTGAGAAAGACAGACGCTAATTAATGCTCGCTG 120  
Db 21 AlathervalasplystrilleGluLeuMetGluGlnlaIaIaArgAsnAsnlaIaArgLeu 40  
QY 121 ATGCGCTTTCCGGAACCTTGATTCAGGCTACCCATGCTTTCTTGCGCTTACTCACA 180  
Db 41 IlelaIaheProglutntrpIleProglYtyrProtrpHeuLeuTrrpleuAspSerPro 60  
QY 181 GCATGGCAATGCAATTGTGACGCAATACCATGAGAACTCATTTGAGAGTGGCCCT 240  
Db 61 AlatrrpIaMetGlnpHevalArgGlnTyrhIsGlnsSerLeuGluLeuAspGlyPro 80  
QY 241 CAAGCTAAGCGCATTTTCAGATGACGCAAGCGGTTGGGAATCATGTCACCTGGGGATG 300  
Db 81 GlnlaIaYsargIleSerAspAlaIaIaYsArgLeuGlyIleMetValThrLeuGlyMet 100  
QY 301 AGTGAACGGGTGCGTGGGACCCCTTACATCAGTCAGTGGTTCATAGGGGATTAATGGTAC 360  
Db 101 SerGlntrgvalGlyGlyThrLeuTyrilleSerGlntrpHeilleGlyAspAsnGlyAsp 120  
QY 361 ACCATTGGGGCCCGCGCAAAAGTTGAAACCTACTTTGTTGAAACGTAACCTTTGTTGGGAA 420  
Db 121 ThrilleGlyAlaIaYsArgLeuYsLeuYsProthrPhevalGlntrgThrLeuPheGlyGlu 140  
QY 421 GGGGATGGTTTCAGCTAGCGGTTTTCAGAGCGTCTGTGGAAAGCGTGGTGGCTTATGC 480  
Db 141 GlyAspGlySerSerleuAlaIaValPheGlntrhSerValGlyArgLeuGlyGlyLeuGly 160  
QY 481 TGTGGGAGCACCTTCAACCCCTTAACAAATACGCTTTGTATGCAAAATGAAGAAGATT 540  
Db 161 CysTrpGlnHisLeuGlnProleuThrYsTrYsTrYsTrYsTrYsTrYsTrYsTrYsTrYs 180  
QY 541 CATTTGGGCGCTGGCGGAGCTTACGCTTATCCCTAATGGCGGAAAGCCCTGGGGCT 600  
Db 181 HisCysAlaIa 200  
QY 601 GATGTCAATGTAGCGGCTCTGCAATCTATGCGCTTGAAGGCAATGCTTCTGATCAAG 660  
Db 201 AspValAsnValaIaIaIaSerArgIleTyrAlaValaIaGlnGlyGlnCysPheValLeuAla 220  
QY 661 TCGTGTGGCTCGTTTCAATCCATGATCGATATGCTTTGTATCAGATGAAGAAAGAT 720  
Db 221 SerCysAlaIa 240  
QY 721 GCGTGTGCTTGGCTGCTGTGACACTCAGCTATCATAGGGCCGATGAGTGTGACTTG 780  
Db 241 AlathleuLeuAlaIaGlyGlyGlyHisSerArgIleilleGlyProAspGlyGlyAspLeu 260  
QY 781 GTGCGGCTCTTGGCCGAAATGAAGAGGGTATCTCTACGCAACCTTGATCTGGAAGTA 840  
Db 261 ValaIaProleuAlaIaGlnsGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGlnGln 280  
QY 841 CGCATCTTGTGTAATAATGGCGGAGACCTGCTGGTCAATTTCCCGCCGACGATTAATCT 900  
Db 281 ArgIleleuAlaIaYsMetAlaIa 300  
QY 901 CGCTTGCTAATAGATGAGCCCTAATTAATTAACCGGTAGTTGAATTAATGAAGTATCTTGT 960  
Db 301 ArgLeuLeuIleAspArgSerProYsLeuProValaIaGlnIleGlnGlyAspLeuArg 320  
QY 961 CCTTAACGCTTTGGGTAAGCGTCTGAGACGGGTGGCGCAACTGAGAAATTT 1011  
Db 321 ProTyrAlaIa 337

RESULT 2  
US-10-146-772-386

Sequence 386, Application US/10146772  
Publication No. US20030124698A1  
GENERAL INFORMATION:  
APPLICANT: Short, Jay  
APPLICANT: Weiner, David  
APPLICANT: Chaplin, Jennifer  
APPLICANT: Chl, Ellen  
APPLICANT: Milan, Allen  
APPLICANT: Desantis, Grace  
APPLICANT: Madden, Mark  
APPLICANT: Burk, Mark  
TITLE OF INVENTION: Nitrlaees  
FILE REFERENCE: Docket No. US20030124698A1 DIV-013US  
CURRENT APPLICATION NUMBER: US/10/146,772  
CURRENT FILING DATE: 2002-05-15  
PRIOR APPLICATION NUMBER: US 60/309,006  
PRIOR FILING DATE: 2001-07-30  
PRIOR APPLICATION NUMBER: US 60/351,336  
PRIOR FILING DATE: 2002-01-22  
PRIOR APPLICATION NUMBER: US 60/300,189  
PRIOR FILING DATE: 2001-06-21  
PRIOR APPLICATION NUMBER: US 09/751,299  
PRIOR FILING DATE: 2000-12-28  
PRIOR APPLICATION NUMBER: US 60/254,414  
PRIOR FILING DATE: 2000-12-07  
PRIOR APPLICATION NUMBER: US 60/173,609  
PRIOR FILING DATE: 1999-12-29  
NUMBER OF SEQ ID NOS: 386  
SOFTWARE: PatsSeq for Windows Version 4.0  
SEQ ID NO 386  
LENGTH: 337  
TYPE: PRT  
ORGANISM: Unknown  
FEATURE:  
OTHER INFORMATION: Obtained from an environmental sample  
US-10-146-772-386  
Alignment Scores:  
Pred. No.: 5.4e-177 Length: 337  
Score: 1761.00 Matches: 336  
Percent Similarity: 100.0% Conservative: 1  
Best Local Similarity: 99.7% Mismatches: 0  
Query Match: 96.4% Gaps: 0  
DB: 4 Indels: 0  
US-09-751-299-3 (1-1014) x US-10-146-772-386 (1-337)  
QY 1 ATGAAGAAGCTATCAAGGTGCGCTGCTGCAAGCCGCCGATCTACATGATTTGGAG 60  
Db 1 MetylsglnatlellysvalaIaCysvalGlnlaIaIaProileTyrMetAspLeuGlu 20  
QY 61 GCGACGGTGGCAAAACATTGATTTGAGAAAGACAGACGCTAATTAATGCTCGCTG 120  
Db 21 AlathervalasplystrilleGluLeuMetGluGlnlaIaIaArgAsnAsnlaIaArgLeu 40  
QY 121 ATGCGCTTTCCGGAACCTTGATTCAGGCTACCCATGCTTTCTTGCGCTTACTCACA 180  
Db 41 IlelaIaheProglutntrpIleProglYtyrProtrpHeuLeuTrrpleuAspSerPro 60  
QY 181 GCATGGCAATGCAATTGTGACGCAATACCATGAGAACTCATTTGAGAGTGGCCCT 240  
Db 61 AlatrrpIaMetGlnpHevalArgGlnTyrhIsGlnsSerLeuGluLeuAspGlyPro 80  
QY 241 CAAGCTAAGCGCATTTTCAGATGACGCAAGCGGTTGGGAATCATGTCACCTGGGGATG 300  
Db 81 GlnlaIaYsargIleSerAspAlaIaIaYsArgLeuGlyIleMetValThrLeuGlyMet 100  
QY 301 AGTGAACGGGTGCGTGGGACCCCTTACATCAGTCAGTGGTTCATAGGGGATTAATGGTAC 360  
Db 101 SerGlntrgvalGlyGlyThrLeuTyrilleSerGlntrpHeilleGlyAspAsnGlyAsp 120  
QY 361 ACCATTGGGGCCCGCGCAAAAGTTGAAACCTACTTTGTTGAAACGTAACCTTTGTTGGGAA 420



```
Db 261 ValAlaProLeuAlaGluAsnGluGluGlyIleLeuYrAlaAsnLeuAspProGlyVal 280
Qy 841 CGCATCTTGTCTAAATGCGCGAGACCCCTGCTGTCATTATTCCTCCGACATTACT 900
Db 281 ArgIleLeuAlaIuSMeAlaAlaAspProAlaGlyHisrYserArgProAspIleThr 300
Qy 901 CGCTTGCTAATAGATCGCAGCCCTTAATTACCGGTAGTTGAAATTGAAGTGTCTTCT 960
Db 301 ArgLeuLeuIleAspArgSerProLysLeuProValAlaGluIleGluGlyAspLeuArg 320
Qy 961 CCTTACGCTTGGGTAAAGCGCTGAGACGGGTGCGACGCAAGCAAAATT 1011
Db 321 ProTyrAlaLeuGlyLysAlaSerGluThrGlyAlaGlnLeuGluGluIle 337

RESULT 4
US-10-440-523-386
; Sequence 386, Application US/10440523
; Publication No. US20040014195A1
; GENERAL INFORMATION:
; APPLICANT: Short, Jay
; APPLICANT: Weiner, David
; APPLICANT: Chaplin, Jennifer
; APPLICANT: Chi, Ellen
; APPLICANT: Milan, Aileen
; APPLICANT: Desantis, Grace
; APPLICANT: Madden, Mark
; APPLICANT: Burk, Mark
; TITLE OF INVENTION: Nucleases
; FILE REFERENCE: Docket No. US20040014195A1 DIV-013US
; CURRENT FILING DATE: 2003-05-15
; PRIOR APPLICATION NUMBER: US/10/146,772
; PRIOR FILING DATE: 2002-05-15
; PRIOR APPLICATION NUMBER: US 60/309,006
; PRIOR FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US 60/351,336
; PRIOR FILING DATE: 2002-01-22
; PRIOR APPLICATION NUMBER: US 60/300,189
; PRIOR FILING DATE: 2001-06-21
; PRIOR APPLICATION NUMBER: US 09/751,299
; PRIOR FILING DATE: 2000-12-28
; PRIOR APPLICATION NUMBER: US 60/254,414
; PRIOR FILING DATE: 2000-12-07
; PRIOR APPLICATION NUMBER: US 60/173,609
; PRIOR FILING DATE: 1999-12-29
; NUMBER OF SEQ ID NOS: 386
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 386
; LENGTH: 337
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Obtained from an environmental sample
US-10-440-523-386

Alignment Scores:
Pred. No.: 5,4e-177 Length: 337
Score: 1761.00 Matches: 336
Percent Similarity: 100.0% Conservative: 1
Best Local Similarity: 99.7% Mismatches: 0
Query Match: 96.4% Indels: 0
DB: 4 Gaps: 0

US-09-751-299-3 (1-1014) x US-10-440-523-386 (1-337)
Qy 1 ATGAAAGAGCTATCAAGTCGCTGCAAGCCGCCGATCTACATGATTGAG 60
Db 1 MetLysGluAlaIleLeuValAlaCysValAlaGlnAlaIleProIleTyrMetAspLeuLys 20
Qy 61 GCGAGCGTGGACAAACCATTTGATGTGAGAAAGACGACGTAATTAAGTCTGCTG 120
Db 21 AlaThrValAspLysThrIleGluMetGluGluAlaAlaArgAsnAsnAlaArgLeu 40
```

```
Qy 121 ATGCGCTTTCCGAAACTTGATTCAGGCTTACCCATGCTTTCTTGGCTTGACTACCA 180
Db 41 IlAlaIleProIleuThrIleProGlyTyrProIlePheLeuThrLeuAspSerPro 60
Qy 181 GCATGGCGAATGCAATTGTAGCCCAATACCAAGAACTCATTTGAGTTGATGGCCCT 240
Db 61 AlaThrPalaMetGlnPheValArgGlnTyrHisGluAsnSerLeuGluLeuAspGlyPro 80
Qy 241 CAAGCTAAGCGCATTTGATGATGACGCCAGCGGTTGGGAATCATGTGACCTTGGGGATG 300
Db 81 GluAlaLysArgIleSerAspAlaAlaLysArgLeuGlyIleMetValIleLeuGlyMet 100
Qy 301 AGTGAACGGGTGGGTGGGACCCCTTTCATGAGTCAGTGCTTCAATAGCGCAATATGGTAC 360
Db 101 SerGluArgValGlyGlyThrLeuTyrIleSerGlnThrPheIleGlyAspAsnGlyAsp 120
Qy 361 ACCATTGGGCGCGCGCAAGTTGAACCTTACTTTTGTGAAGCTATTGTTGCGGGAA 420
Db 121 ThrIleGlyAlaIleArgArgLysLeuLysProThrPheValGluArgThrLeuPheGlyGlu 140
Qy 421 GGGGATGCTTCATGCTAGCGGCTTTTTCGAGACGCTCTTGGAAAGCTGGGTGCTTATGC 480
Db 141 GlyAspGlySerSerLeuAlaValPheGluThrSerValGlyArgLeuGlyGlyLeuCys 160
Qy 481 TGTGGGAGACCTTCACACCGCTTAACAAATAACGCTTGTATGACAAATGAAGAGATT 540
Db 161 CysThrGluThrLeuGlnProLeuThrLysThrAlaLeuTyrAlaGlnAsnGluGluIle 180
Qy 541 CATGTGCGGCTTGGCGGACCTTTAGCCTTATCTTAATGCGGGAAAGCCCTGGGGCT 600
Db 181 HisCysAlaIleAlaIleTrpProSerPheSerLeuTyrProAsnAlaAlaLysAlaLeuGlyPro 200
Qy 601 GATGTCAATGTAGGGGCTCTCGAATCTATGCGCTTGAAGGCAATGCTTGTACTGCG 660
Db 201 AspValAsnValAlaAlaSerArgIleTyrAlaValGluGlyGlnCysPheValLeuAla 220
Qy 661 TCGGTGCGCTCGTTTCAACAATCCATGATGATGATGCTTTGTACAGATGACGAAAGCAT 720
Db 221 SerCysAlaLeuValSerGlnSerMetIleAspMetLeuCysThrThrAspArgLysHis 240
Qy 721 GCGTGTCTTGTGCTGTGTGTGACACTACAGTATCATAGGGCTGATGTGTGACTTG 780
Db 241 AlaLeuLeuLeuAlaGlyGlyGlyHisSerArgIleIleGlyProAspArgLysAspLeu 260
Qy 781 GTCCGCGCTTGGCGGAAATGAAGAGGTATCTCAAGCAAACTTGAATCCGAGGTA 840
Db 261 ValAlaProLeuAlaGluAsnGluGlyIleLeuTyrAlaAsnLeuAspProGlyVal 280
Qy 841 CGCATCTTGTCTAAATGCGCGAGACCCCTGCTGTCATTATTCCTCCGACATTACT 900
Db 281 ArgIleLeuAlaLysMeAlaAlaAspProAlaGlyHisrYserArgProAspIleThr 300
Qy 901 CGCTTGCTAATAGATCGCAGCCCTTAATTACCGGTAGTTGAAATTGAAGTGTCTTCT 960
Db 301 ArgLeuLeuIleAspArgSerProLysLeuProValAlaGluIleGluGlyAspLeuArg 320
Qy 961 CCTTACGCTTGGGTAAAGCGCTGAGACGGGTGCGACGCAAGCAAAATT 1011
Db 321 ProTyrAlaLeuGlyLysAlaSerGluThrGlyAlaGlnLeuGluGluIle 337

RESULT 5
US-10-440-503-386
; Sequence 386, Application US/10440503
; Publication No. US20040038419A1
; GENERAL INFORMATION:
; APPLICANT: Weiner, David Paul
; APPLICANT: Chaplin, Jennifer Ann
; APPLICANT: Chi, Ellen
; APPLICANT: Milan, Aileen
; APPLICANT: Desantis, Grace
; APPLICANT: Burk, Mark J.
; APPLICANT: McQuaid, Jeffrey
```



APPLICANT: Stege, Justin  
TITLE OF INVENTION: ASSAYS AND KITS FOR DETECTING THE PRESENCE OF  
FILE OF INVENTION: NITRILES AND/OR CYANIDE  
FILE REFERENCE: 09010-900001  
CURRENT APPLICATION NUMBER: US/10/440,503  
PRIOR FILING DATE: 2003-05-15  
PRIOR APPLICATION NUMBER: US 60/380,737  
NUMBER OF SEQ ID NOS: 386  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 386  
LENGTH: 337  
TYPE: PRT  
ORGANISM: Unknown  
FEATURE:  
OTHER INFORMATION: Obtained from an environmental sample  
US-10-440-503-386

Alignment Scores:  
Pred. No.: 5.4e-177 Length: 337  
Score: 1761.00 Matches: 336  
Percent Similarity: 100.0% Conservative: 1  
Best Local Similarity: 99.7% Mismatches: 0  
Query Match: 96.4% Indels: 0  
DB: 4 Gaps: 0

US-09-751-299-3 (1-1014) x US-10-440-503-386 (1-337)

QY 1 ATGAAAGAGCTATCAAGGTGCGCTGCGTGAAGCGCCCGATCTACATGATTTGGAG 60  
DB 1 Metysglualalelyvalalacysvalaiglnalaiprolletrymetaspneuyls 20

QY 61 GCGACGGTGAACAAACCATTTGATGATGGAAGACAGACAGTAAATAGTCTGTG 120  
DB 21 Alathrvalasplysthrilegluleumetgluglnalaargasnaasnaalargneu 40

QY 121 ATGCGCTTCGGGAAACTTGATTCAGGCTACCCATGATGCTTGGCTTGAATCAGCA 180  
DB 41 Ilelaahpethroglnthrtprileprogllytrprotrpneutrpneuaspserrpro 60

QY 181 GCATGGGCAATGCAATTTGTAAGCCCAATACCATGGAAGAACTGATGATGGCCCT 240  
DB 61 Alathrplametglnphevalarglntrhrisgluanserleuglnleuaspolypro 80

QY 241 CAAGCTAAGCCCATTTGATGATGACGACGCGGTTGGGAATCATGCTGACCTGGGGATG 300  
DB 81 Glnalalysargylleseraspalaalalysargleuglyllemetvalthleuglymet 100

QY 301 AGTGAACGGGTGGGCGACCCCTTACATGATGATGATGATGATGATGATGATGATGATG 360  
DB 101 Serglunargvalaiglylthleuylrileserglntrpneilleglyaspasnglyasp 120

QY 361 ACCATTTGGGCGCGCGGCAAAAGTTGAACCTACTTTTGTGAAGCTATTTGGCGGAA 420  
DB 121 Thrilleglialaargarglyleuylserprothrphevalaiglnargthleupheglu 140

QY 421 GGGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 480  
DB 141 Glyaspgllyserleuvalaiphelglnthrservalalargleuglyglyleucys 160

QY 481 TGTGGGAGGACCTTCAACCGCTAAACAAATACGCTTGTATGACAAATGAAGAGATT 540  
DB 161 Cyslrpqlunhlsleuglnpneutrhysrlyrthleuylrthleuylrthleuylrthleuyl 180

QY 541 CATTTGGGCGCTTGGCGGCTTTAGCTTTATCTTAATGAGCGAAAGCCCTGGGCGCT 600  
DB 181 Hisysal 200

QY 601 GATGCTAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 660  
DB 201 Aspvalaenval 220

QY 661 TCGGTGTGCGCTGTTTCAATCAATCAATCAATCAATCAATCAATCAATCAATCAATCA 720

DB 221 SerCysalaleuValserGlnSerMetileaspmetleucystrnaspasglulyshts 240

QY 721 GCGTTGCTTCGGTGGTGGTGGACACTGATGATGATGATGATGATGATGATGATGATGATG 780

DB 241 Alaleuenuleualaglyglylyhlseserxglllellegllypneaspasgllyaspneu 260

QY 781 GTCCGCGCTTTCGGGAAATGAAGGGGTATTCATGCAAGAACTGATGATGATGATGATG 840

DB 261 ValalalProleualaglnasngluclyllyleuylrthleuylrthleuylrthleuyl 280

QY 841 GCGATCTTGTAAATGCGCGGACAGCCCTGCTGATTAATTCCTCCGCGGACATTAAT 900

DB 281 ArgilleuualalysemetlialalaspProalaglyhstryserazrgProaspillethr 300

QY 901 CGCTTGTCTAATGATCGGACGCTTAATTAATTCGGTATGTAATTAATGATGATGATG 960

DB 301 ArgleuenuleualaspasgserProlyseuProvalalaglnllyleuglyaspneuarg 320

QY 961 CCTTACGCTTGGGTTAAAGGCTGAGACGGGTGCGGCAACTGGAAGAAATT 1011

DB 321 Protyrthleuuglylysalaserglnthrtglvalaglnleuuglulile 337

RESULT 6  
US-10-461-925-386  
Sequence 386, Application US/10461925  
Publication No. US20040053378A1  
GENERAL INFORMATION:  
APPLICANT: Mark J. Burk  
APPLICANT: Desantis, Grace  
APPLICANT: Morgan, Brian  
APPLICANT: Zhu, Zoulin  
TITLE OF INVENTION: PROCESSES FOR MAKING (R)-ETHYL 4-CYANO-3-HYDROXYBUTYRIC ACID  
FILE REFERENCE: 09010-270001  
CURRENT APPLICATION NUMBER: US/10/461,925  
PRIOR FILING DATE: 2003-06-13  
PRIOR APPLICATION NUMBER: US 60/389,317  
PRIOR FILING DATE: 2002-06-13  
PRIOR APPLICATION NUMBER: US 60/392,944  
PRIOR FILING DATE: 2002-06-28  
NUMBER OF SEQ ID NOS: 386  
SOFTWARE: FastSeq for Windows Version 4.0  
SEQ ID NO 386  
LENGTH: 337  
TYPE: PRT  
ORGANISM: Unknown  
FEATURE:  
OTHER INFORMATION: Obtained from an environmental sample  
US-10-461-925-386

Alignment Scores:  
Pred. No.: 5.4e-177 Length: 337  
Score: 1761.00 Matches: 336  
Percent Similarity: 100.0% Conservative: 1  
Best Local Similarity: 99.7% Mismatches: 0  
Query Match: 96.4% Indels: 0  
DB: 4 Gaps: 0

US-09-751-299-3 (1-1014) x US-10-461-925-386 (1-337)

QY 1 ATGAAAGAGCTATCAAGGTGCGCTGCGTGAAGCGCCCGATCTACATGATTTGGAG 60  
DB 1 Metysglualalelyvalalacysvalaiglnalaiprolletrymetaspneuyls 20

QY 61 GCGACGGTGAACAAACCATTTGATGATGGAAGACAGACAGTAAATAGTCTGTG 120  
DB 21 Alathrvalasplysthrilegluleumetgluglnalaargasnaasnaalargneu 40

QY 121 ATGCGCTTCGGGAAACTTGATTCAGGCTACCCATGATGCTTGGCTTGAATCAGCA 180  
DB 41 Ilelaahpethroglnthrtprileprogllytrprotrpneutrpneuaspserrpro 60

QY 181 GCATGGGCAATGCAATTTGTAAGCCCAATACCATGGAAGAACTGATGATGGCCCT 240

```
Db      ||| 61 Alatrpalamctglnpnevallarglnttyrhiseulunsenrleugluleaspelgylpro 80
Qy      ||| 241 CAAGCTAAGCGCATTTGAGATGACGCAAGCGGTTGGGAATCATGTCACCTTGGGATG 300
Db      ||| 81 Glnhlaalysarglleseraspalaalalyargleuglyilemetvalthrleuglymet 100
Qy      ||| 301 AGTGAAGGGTGGTGGGACCCCTTACATGATGATGAGTTCATAGGCGATTAATGATGAC 360
Db      ||| 101 Serlunargvalglgylthrleutyrlleserlntppheillegllyaspasnlglyasp 120
Qy      ||| 361 ACCATTGGGGCCCGCGCAAAAGTTGAAACCTTCTTTGTTGAACGTAACCTTTTGGCGAA 420
Db      ||| 121 Thrleeglyalaagaaglyleuleyprothrphelvalgluaagthlrlephelgylglu 140
Qy      ||| 421 GGGGATGGTTCATGCTGACCGGTTTGGAGAGCTGTGGAGAGCTGGGCTTATGC 480
Db      ||| 141 Glyaspglserleuvalaiphelglnhrservalgllyargleuglyleucys 160
Qy      ||| 481 TGTGGGAGCACTTCAACCGCTAACAATAACGCTTGTATGACAAATGAAGAGATT 540
Db      ||| 161 Cysrlpglnhlsleuglnproleuthrlystyralauleuryralaaglnasnglunlle 180
Qy      ||| 541 CATGTGCGGCTTGGCGAGCTTTAGCCTTTATCTTAATGCGGCGAAAGCCCTGGGCT 600
Db      ||| 181 Hiscysalaalatrproserpheserleutyrrhoasnaalalalyalaleuglypro 200
Qy      ||| 601 GATGTCAATGTAGCGGCTCTCGAATCTATGCCCTTGAAGGGCAATGTTGTACTAGC 660
Db      ||| 201 Aspvalasnvalaalaaserarglleyrvalaaglnuglyncysphevalleuula 220
Qy      ||| 661 TCGGTGGGCTCGTTTCAATCATGATGATGATGATGATGATGATGATGATGATGATGAT 720
Db      ||| 221 Serlysalauleuvalserglnsermetlleasphetleucysthrhaspbglnuylhis 240
Qy      ||| 721 GCGTGTCTTCTGCTGTGTGTGACACTCACGTATCATATAGAGGCTGATGTGTGACTTG 780
Db      ||| 241 Alaleuenuleuualaglygllyhisserarglleileglproaspglygllyaspneu 260
Qy      ||| 781 GTGCGGCTCTTGGCGAAATGAAGAAGGATTTCTTACGCAACCTTGAATCTGGAGTA 840
Db      ||| 261 Valalaproleuualaglnasnglunllyleuuryralaasnleuasprrgilyal 280
Qy      ||| 841 CGCATCTTGTCTAATAATGGCGGAGACCGCTGCTGCTATTCGCCGCGCGACATTAAT 900
Db      ||| 281 Argilleuualalysemetalaalasprrhlaaglyhlsrlyseralyspproaspillethr 300
Qy      ||| 901 CGCTTGCTATATGATCGCAGCCCTAATTAATCCGCTAGTTGAATTTGAAGTGAATCTTGC 960
Db      ||| 301 Argleuenuleuasphargserprolyleuuprovalvalgluilegllyaspneuarg 320
Qy      ||| 961 CCTTACGCTTTGGTGAAGCGCTCGAGCGGGTGGGCAACTCGAAGAAATT 1011
Db      ||| 321 Protyralaleuuglylvalaserclunhrghylaglnleuuglunlle 337

RESULT 7
US-10-146-772-48
; Sequence 48, Application US/10146772
; Publication No. US20030124698A1
; GENERAL INFORMATION:
; APPLICANT: Short, Jay
; APPLICANT: Weiner, David
; APPLICANT: Chaplin, Jennifer
; APPLICANT: Chi, Ellen
; APPLICANT: Milan, Aileen
; APPLICANT: Desantis, Grace
; APPLICANT: Maden, Mark
; APPLICANT: Burk, Mark
; TITLE OF INVENTION: Nitrlases
; FILE REFERENCE: Docket No. US20030124698A1 DIV-013US
; CURRENT APPLICATION NUMBER: US/10/146,772
; PRIOR APPLICATION NUMBER: US 60/309,006
```

```
;; PRIOR FILING DATE: 2001-07-30
;; PRIOR APPLICATION NUMBER: US 60/351,336
;; PRIOR FILING DATE: 2002-01-22
;; PRIOR APPLICATION NUMBER: US 60/300,189
;; PRIOR FILING DATE: 2001-06-21
;; PRIOR APPLICATION NUMBER: US 09/751,299
;; PRIOR FILING DATE: 2000-12-28
;; PRIOR APPLICATION NUMBER: US 60/254,414
;; PRIOR FILING DATE: 2000-12-07
;; PRIOR APPLICATION NUMBER: US 60/173,609
;; PRIOR FILING DATE: 1999-12-29
;; NUMBER OF SEQ ID NOS: 386
;; SOFTWARE: PasterSeq for Windows Version 4.0
;; SEQ ID NO 48
;; LENGTH: 337
;; TYPE: PRT
;; ORGANISM: Unknown
;; FEATURE:
;; OTHER INFORMATION: Obtained from an environmental sample
US-10-146-772-48

Alignment Scores:
Pred. No.: 2,56e-132 Length: 337
Score: 1339.00 Matches: 252
Percent Similarity: 84.2% Conservative: 30
Best local Similarity: 75.2% Mismatches: 47
Query Match: 73.3% Indels: 6
DB: 4 Gaps: 1

US-09-751-299-3 (1-1014) x US-10-146-772-48 (1-337)
Qy      ||| 1 ATGAAGAAGCTATCAAGTGCCTGCGTCAAGCCGCCGATCATGATTTGAG 60
Db      ||| 1 VallyselunlailelysvlaiaCyvalaGlnhlaalatrprovalphleuaspneuasp 20
Qy      ||| 61 GCGACGGTGAACAACCATTTGATGATGAGAAAGACACGACGTAATTAATGCTGCTG 120
Db      ||| 21 Alatrvalaspysrthrvalaaleuileglunlaalatrvasnglylaalargleu 40
Qy      ||| 121 ATGCGCTTTCCGGAATTTGATTCAGGCTACCAAGTTCCTTGGCTTGACTGACA 180
Db      ||| 41 llelaheprrglnuhtrrpilleprrgilyrtrprrhlepneuuprleuaspserpro 60
Qy      ||| 181 GCATGGGCAATGCAATTTGTACGCAATATACATGAGAACTCATTTGAGTGTGAGCCCT 240
Db      ||| 61 Alatrpglymetglnpnevallargargtyrhiseulunsenrleuvalleuaspserpro 80
Qy      ||| 241 CAAGCTAAGCGCATTTGAGATGACGCAAGCGGTTGGGAATCATGTCACCTTGGGATG 300
Db      ||| 81 Glnhlaalysarglleserlunlaalaglnhraglaaglylleyrvalaaleuglytyr 100
Qy      ||| 301 AGTGAAGGGTGGTGGGACCCCTTACATGATGATGAGTTCATAGGCGATTAATGATGAC 360
Db      ||| 101 Serlunargvalsergylthrleutyrmecglylntppheulleasphslysglyglu 120
Qy      ||| 361 ACCATTGGGGCCCGCGCAAAAGTTGAAACCTTCTTTGTTGAACGTAACCTTTTGGCGAA 420
Db      ||| 121 Thrleaglyleuargaglyleuleyprothrhrhlsvalgluaagthlrlephelgylglu 140
Qy      ||| 421 GGGGATGGTTCATGCTGACCGGTTTGGAGAGCTGTGGAGAGCTGGGCTTATGC 480
Db      ||| 141 Glyaspglserleuaserthrphespphrproleuuglyvalleuuglyleucys 160
Qy      ||| 481 TGTGGGAGCACTTCAACCGCTAACAATAACGCTTGTATGACAAATGAAGAGATT 540
Db      ||| 161 Cysrlpglnhlsleuglnproleuserlystyralauleuryralaaglnasnglunlle 180
Qy      ||| 541 CATGTGCGGCTTGGCGAGCTTTAGCCTTTATCTTAATGCGGCGAAAGCCCTGGGCT 600
Db      ||| 181 Hiscysalaalatrproserpheserilletyrarglnlaalthrghylvalleuglypro 200
Qy      ||| 601 GATGTCAATGTAGCGGCTCTCGAATCTATGCCCTTGAAGGGCAATGTTGTACTAGC 660
;;
```





```

; PRIOR FILING DATE: 1999-12-29
; NUMBER OF SEQ ID NOS: 386
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 58
; LENGTH: 337
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Obtained from an environmental sample
US-10-241-742-58

Alignment Scores:
Pred. No.: 2,566-132 Length: 337
Score: 1339.00 Matches: 252
Percent Similarity: 84.2% Conservative: 30
Best Local Similarity: 75.2% Mismatches: 47
Query Match: 73.3% Indels: 6
DB: Gaps: 1

US-09-751-299-3 (1-1014) x US-10-241-742-58 (1-337)

QY 1 ATGAAAGAGCTATCAAGTCCCTGCGTCAAGCCGCCGATCTACATGATTTGGAG 60
   |||||
DB 1 Vallysglualailelysvaliacysvalglnalalaprovalpheleuaspseu 20
   |||||

QY 61 GCGACGGTGGACAAACCATGAGTTGATGGAAGAAAGACGTAAATAGCTGCTG 120
   |||||
DB 21 Alatrvalasplvthrvalalaleuilegluglnalalargasnilyalargleu 40
   |||||

QY 121 ATGCGCTTCCGGAACCTTGATTCAGGCTCCATGCTTCTTGCTGACTCACA 180
   |||||
DB 41 lleaapherprogluthrtrpleproglyrprotrptheleuttrpleuaspse 60
   |||||

QY 181 CGATGGGCATGCAATTTGTACGCCAATACATGAGACTCATTTGAGTTGAGCCCT 240
   |||||
DB 61 Alatrpglymetglnphevalargarglyrhisgluasrleuvalleuaspse 80
   |||||

QY 241 CAAGCTAAGCGGATTCAGATGACGAGCGGTTGGGAATGATGTCACCTGGGAGT 300
   |||||
DB 81 Glnalalyarglleserlualalaglnarglalylyrvalalaleucllyr 100
   |||||

QY 301 AGTGAACGGGTGCGTGGCACCCCTTACATCAGTCAAGTGTGATGAGCGATATG 360
   |||||
DB 101 Sergluargvalasercllythrleuylrmetglylntprleuileaspaspl 120
   |||||

QY 361 ACCATTTGGGGCCGGGGAAGTTGAAACCTACTTTGTTGAAAGTATTTGTCGCG 420
   |||||
DB 121 ThrllaglyleuargarglyseuylsProthrhisvalgluargthreuphe 140
   |||||

QY 421 GGGGATGCTTCACTGACCGGTTTCCAGACGTCTGTGGAAGCTGGGTGCTTAT 480
   |||||
DB 141 GllaypbglyserseuilerthreapthrProleuglyvalleucllyleu 160
   |||||

QY 481 TGTGGGAGACCTTCAACCGCTAAACAAATAGCTTTGTATGCAAAATAGAGAT 540
   |||||
DB 161 CyslrpglnhisleuclnProleuserlystryalaleuylrhisgluhsn 180
   |||||

QY 541 CATTTGCGGCTTGGCGGAGCTTTAGCTTTATCTTAATGCGGAAAGCCCTGGG 600
   |||||
DB 181 HlsphaelaalatrProserpheserllyrarglnalathrglvalleucl 200
   |||||

QY 601 GATGTCATGTAGCGGCTCTGGAATCTATGCCCTTGAAGGCGCATGCTTGTACT 660
   |||||
DB 201 Gluvalasvalalalaserarglyrlyalavalegluglylncysphevalle 220
   |||||

QY 661 TGTGTGCGCTGCTTCAATCATCATGATCATGATGCTTTGACAGATGACAAAG 720
   |||||
DB 221 SerCybalaleuvalaserProglumetclleglumetleuCythrpscl 240
   |||||

QY 721 GCGTTCCTTCTGCGTGGTGGACACTCAGATTCATATAGGCGCTGATGCTGAT 780
   |||||
DB 241 SerleuclnaglailaglyllyrserargllelellyProaspolyserasp 260
   |||||

```

```

QY 781 GTGCGCCTCTGCGGAAATGAGAGGATATTTCTATCGCAAACTTGATCTGAGTA 840
   |||||
DB 261 AlargProleucllylualsnclugluclylleuylrhalatrtheuaspPro 280
   |||||

QY 841 CGCATCTTGTCTAAATGGCGGACAGACCTGCTGCTATTAATCCGCTCCGACAT 900
   |||||
DB 281 ArglyrlyalalythrllalalaspProalaglylstryserargProaspval 300
   |||||

QY 901 CGCTTGCAATAGATCGGAGCCCTAAATTACCGGTAGTTGAATGAAGGATCTG 960
   |||||
DB 301 ArgleuclleuileasrargseralasnclnProvalvalgluvalgllyarg 319
   |||||

QY 961 CTTACGCTTGGGTAAAGCCTGTGAGCGGCTGCGCACTCGAA 1005
   |||||
DB 320 -----Proalaserlalnclnlypnegluvalglu 329
   |||||

RESULT 11
US-10-440-523-48
; Sequence 48, Application US/10440523
; Publication No. US20040014195A1
; GENERAL INFORMATION:
; APPLICANT: Short, Jay
; APPLICANT: Weiner, David
; APPLICANT: Chaplin, Jennifer
; APPLICANT: Chi, Ellen
; APPLICANT: Milan, Aileen
; APPLICANT: Desantis, Grace
; APPLICANT: Madden, Mark
; APPLICANT: Burk, Mark
; TITLE OF INVENTION: Nitrilases
; FILE REFERENCE: Docket No. US20040014195A1 DIV-013US
; CURRENT APPLICATION NUMBER: US/10/440,523
; PRIOR FILING DATE: 2003-05-15
; PRIOR APPLICATION NUMBER: US/10/146,772
; PRIOR FILING DATE: 2002-05-15
; PRIOR APPLICATION NUMBER: US 60/309,006
; PRIOR FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US 60/351,336
; PRIOR FILING DATE: 2002-01-22
; PRIOR APPLICATION NUMBER: US 60/300,189
; PRIOR FILING DATE: 2001-06-21
; PRIOR APPLICATION NUMBER: US 09/751,299
; PRIOR FILING DATE: 2000-12-28
; PRIOR APPLICATION NUMBER: US 60/254,414
; PRIOR FILING DATE: 2000-12-07
; PRIOR APPLICATION NUMBER: US 60/173,609
; PRIOR FILING DATE: 1999-12-29
; NUMBER OF SEQ ID NOS: 386
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 48
; LENGTH: 337
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Obtained from an environmental sample
US-10-440-523-48

Alignment Scores:
Pred. No.: 2,566-132 Length: 337
Score: 1339.00 Matches: 252
Percent Similarity: 84.2% Conservative: 30
Best Local Similarity: 75.2% Mismatches: 47
Query Match: 73.3% Indels: 6
DB: Gaps: 1

US-09-751-299-3 (1-1014) x US-10-440-523-48 (1-337)

QY 1 ATGAAAGAGCTATCAAGTCCCTGCGTCAAGCCGCCGATCTACATGATTTGGAG 60
   |||||
DB 1 Vallysglualailelysvaliacysvalglnalalaprovalpheleuaspse 20
   |||||

QY 61 GCGACGGTGGACAAACCATGAGTTGATGGAAGAAAGACGTAAATAGCTGCTG 120
   |||||

```

```
Db      21  AlatrValaplysThrValAlaleuileglunlualalargAsnglYAlaArgleu 40
Qy      121  ATCCGCTTCCGGAACCTTGAAATCCAGGCTAACCATGGTTCTTTGGCTTGACTCACCA 180
Db      41  IlealApeProglunThripIleProglYTrProtrPheleutripLeuAspSerPro 60
Qy      181  GCATGGGCAATGCAATTTGTAGCGCAATACCATGGAAGCACTGATGGAGTGGAGCCCT 240
Db      61  AlatrpglYmeGlnPheValalargArgYrhiSeglunSenserleValleuAspSerPro 80
Qy      241  CAAGCTAAGCGCATTTTCAGATGACAGCCAGCGGTTGGGAATCATAGTGCACCTTGGGAG 300
Db      81  GlnAlaySaArgIleSerGlnAlalaglInlrgAlagIlyIleYrValAlaleuGlYTr 100
Qy      301  AGTGAACGGGTCCGATGGGACCCCTTACATGCATGCATGGTTTCATAGGGAATGAAGTAC 360
Db      101  SerGlnaGValSerGlyThrleuYrWecIyGlnTrpleuIleAspAspIyglu 120
Qy      361  ACCATTGGGGCCCGCGGAAAGTTGAACCTACTTTTGTGAAACGTACTTGTTCGGCGAA 420
Db      121  ThrIaGlYleuAlargIyLeuYrSerProthrhISValGlnaArgThrleuPheGlYglu 140
Qy      421  GGGGATGGTTTACCTGCTAGCGGTTTTCAGAGAGCTGTGTTGAAAGGCTGGGCTTATGC 480
Db      141  GlyaSpGlySerSerleuSerThrPheAspThrProleuGlYValleuGlYleuCYs 160
Qy      481  TGTGGAGACACCTTCAACCGCTAACAAATACGCTTGTGTGACAAATGAAGAGATT 540
Db      161  CysTrpGlnhISleuGlnProleuSerYrStYrAlaleuYrAlaGlnhISnGlunIle 180
Qy      541  CATTGTCGGCTTGGCCGAGCTTTAGCCTTATTCCTAATGCGGCAAAACCTTGGGSCCT 600
Db      181  hISpheaIalatrProSerPheSerIleYrArgGlnIalathGlnValleuGlYPro 200
Qy      601  GATTCATATGACGGCCCTTCGATCTATATGCGGTTGAAGGCAATGCTTGCTAGTACG 660
Db      201  GlnValaInhValAlalaserArgIleYrAlaValaGlnGlYGlnCySpheValleuAla 220
Qy      661  TCGTGTGGCTCGTTTCAATCCATTCATGCATATGCTTTGTGACAGATGACAAAGCAT 720
Db      221  SerCySaIaleuValSerProglunWecIleGlunWecIleuCYsThrAspGlnSerIyShIS 240
Qy      721  GCGTTCCTTCTGGCTGGTGGTGAACATCACTCATATGAGGCTCGATGGTGGTACTTG 780
Db      241  SerleuLeuGlnAlagIyglYrYrSerArgIleIleGlYProAspGlySerAspLeu 260
Qy      781  GTCGGCGCTTCCGCAAAATGAAAGGGTATTTCTTACGCAACCTTGATCTGAGAGTA 840
Db      261  AlaargProleuGlYglunSnglunGlYIleleuYrAlaThrleuAspProAlaIa 280
Qy      841  CGCATCTTGTAAATGGCGGCAAGCCCTGCTGCTATTATTCCTCCGACATTAAT 900
Db      281  ArgIleYrAlaYrThrAlaIaAspProAlaISySerArgProAspValThr 300
Qy      901  CGCTTGCTTAATGATCGGAGCCCTTAATTAACGGGTATGGAATGAAGTGAATCTTCT 960
Db      301  ArgleuLeuIleAspArgSerAlaAsnglInProValaGlnValaGlnArgIleUle--- 319
Qy      961  CCTTACGCTTGGGTAAAGCTGTGAGAGCGGTGGCGCACTGCAA 1005
Db      320  -----ProhIaSerIaGlnGlYpHeGlValaGln 329

RESULT 12
US-10-440-523-58
; Sequence 58, Application US/10440523
; Publication No. US20040014195A1
; GENERAL INFORMATION:
; APPLICANT: Short, Jay
; APPLICANT: Weiner, David
; APPLICANT: Chaplin, Jennifer
; APPLICANT: Chi, Ellen
; APPLICANT: Milan, Aileen
; APPLICANT: Desautels, Grace
```

```
APPLICANT: Madden, Mark
APPLICANT: Burk, Mark
TITLE OF INVENTION: Nitrlases
FILE REFERENCE: Docket No. US20040014195A1 DIV-013US
CURRENT APPLICATION NUMBER: US/10/440,523
CURRENT FILING DATE: 2003-05-15
PRIOR APPLICATION NUMBER: US/10/146,772
PRIOR FILING DATE: 2002-05-15
PRIOR APPLICATION NUMBER: US 60/309,006
PRIOR FILING DATE: 2001-07-30
PRIOR APPLICATION NUMBER: US 60/351,336
PRIOR FILING DATE: 2002-01-22
PRIOR APPLICATION NUMBER: US 60/300,189
PRIOR FILING DATE: 2001-06-21
PRIOR APPLICATION NUMBER: US 09/751,299
PRIOR FILING DATE: 2000-12-28
PRIOR APPLICATION NUMBER: US 60/254,414
PRIOR FILING DATE: 2000-12-07
PRIOR APPLICATION NUMBER: US 60/173,609
PRIOR FILING DATE: 1999-12-29
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 58
LENGTH: 337
TYPE: PRT
ORGANISM: Unknown
FEATURE:
OTHER INFORMATION: Obtained from an environmental sample
US-10-440-523-58

Alignment Scores:
Pred. No.: 2,56e-132 Length: 337
Score: 1339.00 Matches: 252
Percent Similarity: 84.2% Conservative: 30
Best Local Similarity: 75.2% Mismatches: 47
Query Match: 73.3% Indels: 6
DB: 4 Gaps: 1

US-09-751-299-3 (1-1014) x US-10-440-523-58 (1-337)
Qy      1  ATGAAGAAGGTATCAAGTGTGCTGCTGCAAGCCCGCATCTACATGATTTGGAG 60
Db      1  ValysGlnAlaIleYrValAlaCyValaGlnAlaIaProValPheLeuAspLeuAsp 20
Qy      61  GCGAGGTGCAAAACCATTTGAAGTGAAGAGACGACGTAATATGCTCGTCTG 120
Db      21  AlatrValaplysThrValAlaleuileglunlualalargAsnglYAlaArgleu 40
Qy      121  ATCCGCTTCCGGAACCTTGAAATCCAGGCTAACCATGGTTCTTTGGCTTGACTCACCA 180
Db      41  IlealApeProglunThripIleProglYTrProtrPheleutripLeuAspSerPro 60
Qy      181  GCATGGGCAATGCAATTTGTAGCGCAATACCATGGAAGCACTGATGGAGTGGAGCCCT 240
Db      61  AlatrpglYmeGlnPheValalargArgYrhiSeglunSenserleValleuAspSerPro 80
Qy      241  CAAGCTAAGCGCATTTTCAGATGACAGCGGTTGGGAATCATAGTGCACCTTGGGAG 300
Db      81  GlnAlaySaArgIleSerGlnAlalaglInlrgAlagIlyIleYrValAlaleuGlYTr 100
Qy      301  AGTGAACGGGTCCGATGGGACCCCTTACATGCATGCATGGTTTCATAGGGAATGAAGTAC 360
Db      101  SerGlnaGValSerGlyThrleuYrWecIyGlnTrpleuIleAspAspIyglu 120
Qy      361  ACCATTGGGGCCCGCGGAAAGTTGAACCTACTTTTGTGAAACGTACTTGTTCGGCGAA 420
Db      121  ThrIaGlYleuAlargIyLeuYrSerProthrhISValGlnaArgThrleuPheGlYglu 140
Qy      421  GGGGATGGTTTACCTGCTAGCGGTTTTCAGAGAGCTGTGTTGAAAGGCTGGGCTTATGC 480
Db      141  GlyaSpGlySerSerleuSerThrPheAspThrProleuGlYValleuGlYleuCYs 160
Qy      481  TGTGGAGACACCTTCAACCGCTAACAAATACGCTTGTGTGACAAATGAAGAGATT 540
```

```

Db      161 CysTPRGlnHisLeuGlnProLeuSerIyStryAlaLeuTyrAlaGlnSngIuGluile 180
Qy      541 CATTTGGCGGCTTTGGCGGAGCTTTAGCTTTATCTTAATGCGGAGAACCTTGGGCGCT 600
Db      181 HisPheAlaAlaIatPrProSerPheSerIleTyrArgGlnAlaThrCgluValIeuGlyPro 200
Qy      601 GATGCAATGTAAGCGGCTCTGGAATCTATGCGGTGAAGGCAATGCTGCTGATAGCGG 660
Db      201 GluValAsnValAlaAlaSerArgIleTyrAlaValGluGlyGlnCyPheValIeuAla 220
Qy      661 TGTGTGCGCTCGTTTCACAAATCCATGATGATGCTTTGTACAGATGACAAAGCAT 720
Db      221 SerCyAlaLeuValSerProGluMetIleGluMetLeuCyethrAspGluSerIyShs 240
Qy      721 GCGTTGCTTCTGAGCTGTGTGTGACACACGATATGAGGCGGTGATGTGTGACTTG 780
Db      241 SerLeuLeuGlnAlaGlyGlyGlyTyrSerArgIleIleGlyProAspGlySerAspLeu 260
Qy      781 GTGCGGCGCTTGGCGGAAATGAAGGGTATTTCTTACGAAACCTTGATCTGGAGTA 840
Db      261 AlaArgProLeuGlyGluSngIuGlyIleLeuTyrAlaThrLeuAspProAlaAla 280
Qy      841 CGCATCTTTGCTAAATGCGGAGACCTGCTGTGCTATTATCCGTCGCGACATTAAT 900
Db      281 ArgIleTyrAlaIySthrAlaAlaAspProAlaGlyHisTyrSerArgProAspValThr 300
Qy      901 GCGTTGCTAATAGATCGAGCGCTTAATTAACCGGTAGTTGAATTAAGATGATCTTCT 960
Db      301 ArgLeuLeuIleAsnArgSerAlaAsnGlnProValIleGlyArgGluIle--- 319
Qy      961 CCTTACGCTTTGGGTAAAGCGTCTGAGACGGGTGGCGCAACTCGAA 1005
Db      320 -----ProAlaSerAlaGlnGlyPheGluValGlu 329

```

```

RESULT 13
US-10-440-503-48
; Sequence 48, Application US/10440503
; Publication No. US20040038419A1
; GENERAL INFORMATION:
; APPLICANT: Weiner, David Paul
; APPLICANT: Chaplin, Jennifer Ann
; APPLICANT: Chi, Ellen
; APPLICANT: Milan, Aileen
; APPLICANT: Desantis, Grace
; APPLICANT: Burk, Mark J.
; APPLICANT: McQuaid, Jeffrey
; APPLICANT: Stege, Justin
; TITLE OF INVENTION: ASSAYS AND KITS FOR DETECTING THE PRESENCE OF
; FILE REFERENCE: 09010-900001
; CURRENT APPLICATION NUMBER: US/10/440,503
; CURRENT FILING DATE: 2003-05-15
; PRIOR APPLICATION NUMBER: US 60/380,737
; PRIOR FILING DATE: 2002-05-15
; NUMBER OF SEQ ID NOS: 386
; SOFTWARE: FASTSQ for Windows Version 4.0
; SEQ ID NO 48
; LENGTH: 337
; TYPE: PRF
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Obtained from an environmental sample
US-10-440-503-48

```

```

Alignment Scores:
Pred. No.: 2,566-1332 Length: 337
Score: 1339.00 Matches: 252
Percent Similarity: 84.2% Conservative: 30
Best Local Similarity: 75.2% Mismatches: 47
Query Match: 73.3% Indels: 6
Db: 4 Gaps: 1

```

```

US-09-751-299-3 (1-1014) x US-10-440-503-48 (1-337)
Qy      1 ATGAAAGAGCGTATTAAGTGGCGCTTGCAGACGCCGCCGATCTAGATGATTTGGAG 60
Db      1 ValIySglnAlaIleValValAlaCyValGlnAlaAlaProValPheLeuAspLeuAsp 20
Qy      61 GCGACGGGTGACAAACCATGATGATGTAAGGAGACGACGTAATTAATGCTGCTG 120
Db      21 AlaThrValAspIySthrValAlaLeuIleGluGlnAlaAlaArgAsnGlyAlaArgLeu 40
Qy      121 ATGCGCTTTCCGGAATCTGGAATTCAGAGCTTACCCAGTGTCTTTGCTGACTACCA 180
Db      41 IleAlaPheProGluThrTrpIleProGlyTyrProIlePheLeuTrpLeuAspSerPro 60
Qy      181 GCATGGGCAATGCAATTTGTACGCCAATACCATGAGAACTCATTTGATGATGAGCCCT 240
Db      61 AlaTrpGlyMetGlnPheValArgArgTyrHisGluAsnSerLeuValLeuAspSerPro 80
Qy      241 CAAGCTAAGCGCATTTGATGACAGCAGCGGCTTGGGAAATCATAGTCAACCTTGGGAGT 300
Db      81 GlnAlaIySArgIleSerGlnAlaAlaGlnArgAlaGlyIleTyrValAlaLeuGlyTyr 100
Qy      301 AGTGAACGGGTGCGGTGACACCTTTATCATGATGATGATGATGATGATGATGATGATG 360
Db      101 SerGlnArgValSerGlyThrLeuTyrMetGlyGlnTrpLeuIleAspAspIySgln 120
Qy      361 ACCATTTGGGCGCGCGGCGGAAAGTTGAACCTTATTTGTTGAACGTAATTTGTTGCGG 420
Db      121 ThrAlaGlyLeuArgArgIySleuIySProThrHisValGluArgThrLeuPheGlyGln 140
Qy      421 GGGGAGTGTATGCTGCTGCTGCGGTTTGCAGACGCTGTGGAAGGCTGGGTGCTTATGC 480
Db      141 GlyAspGlySerSerLeuSerThrPheAspThrProLeuGlyValLeuGlyIySleuCy 160
Qy      481 TGTGGGAGACCTTCAACCGCTAACAAATACGCTTGTATGCAAAATGAAGAT 540
Db      161 CysTPRGlnHisLeuGlnProLeuSerIyStryAlaLeuTyrAlaGlnSngIuGluile 180
Qy      541 CATTTGGCGGCTTTGGCGGAGCTTTAGCTTTATCTTAATGCGGAGAACCTTGGGCGCT 600
Db      181 HisPheAlaAlaIatPrProSerPheSerIleTyrArgGlnAlaThrCgluValIeuGlyPro 200
Qy      601 GATGCAATGTAAGCGGCTCTGGAATCTATGCGGTGAAGGCAATGCTGCTGATAGCGG 660
Db      201 GluValAsnValAlaAlaSerArgIleTyrAlaValGluGlyGlnCyPheValIeuAla 220
Qy      661 TGTGTGCGCTCGTTTCACAAATCCATGATGATGCTTTGTACAGATGACAAAGCAT 720
Db      221 SerCyAlaLeuValSerProGluMetIleGluMetLeuCyethrAspGluSerIyShs 240
Qy      721 GCGTTGCTTCTGAGCTGTGTGTGACACCTGATATGAGGCGGTGATGTGTGACTTG 780
Db      241 SerLeuLeuGlnAlaGlyGlyGlyTyrSerArgIleIleGlyProAspGlySerAspLeu 260
Qy      781 GTGCGGCGCTTGGCGGAAATGAAGGGTATTTCTTACGAAACCTTGATCTGGAGTA 840
Db      261 AlaArgProLeuGlyGluSngIuGlyIleLeuTyrAlaThrLeuAspProAlaAla 280
Qy      841 CGCATCTTTGCTAAATGCGGAGACCTGCTGTGCTATTATCCGTCGCGACATTAAT 900
Db      281 ArgIleTyrAlaIySthrAlaAlaAspProAlaGlyHisTyrSerArgProAspValThr 300
Qy      901 GCGTTGCTAATAGATCGAGCGCTTAATTAACCGGTAGTTGAATTAAGATGATCTTCT 960
Db      301 ArgLeuLeuIleAsnArgSerAlaAsnGlnProValIleGlyArgGluIle--- 319
Qy      961 CCTTACGCTTTGGGTAAAGCGTCTGAGACGGGTGGCGCAACTCGAA 1005
Db      320 -----ProAlaSerAlaGlnGlyPheGluValGlu 329

```

```

RESULT 14
US-10-440-503-58
; Sequence 58, Application US/10440503

```





```

QY 61 GCGACGGTGGACAAACCATTTGATGTGAAGAAGACGACATTAATATGCTGCTG 120
DB 21 AlatThrValAspLysThrValAlaLeuIleGluGluAlaIleArgAsnGlyAlaArgLeu 40
QY 121 ATGCGCTTTCCGGAACCTTGATTCAGGCTAACCCATGGTTTCTTTGGCTTGACTCACCA 180
DB 41 IleAlaPheProGluThrTrpIleProGlyTyrProTrpPheLeuTrpLeuAspSerPro 60
QY 181 GCATGGGCAATGCAATTTGTAGCCCAATACCTGAAGAACTCATTTGAGTTGATGGCCCT 240
DB 61 AlaTrpGlyMetGlnPheValArgTrpHisGluAsnSerLeuValLeuAspSerPro 80
QY 241 CAAGCTAAGCGCATTTACATGCACGACGCGGTGGGAAATCATGTGTCACCTCGGAGATG 300
DB 81 GlnAlaLysArgIleSerGlnAlaAlaGlnArgAlaGlyIleTyrValAlaLeuGlyTyr 100
QY 301 AGTGAACGGGTGGTGGACCCCTTTACATCACTCACTAGTGGTTCATAGCGCATATATGTTAC 360
DB 101 SerGlnArgValSerGlyThrLeuTyrMetGlyGlnTrpLeuIleAspAspLysGlyGlu 120
QY 361 ACCATTTGGGGGCGCGGCAAACTTGAACCTACTTTGTGAACTGTTGTTGCTTGGCGGAA 420
DB 121 ThrAlaGlyLeuArgLysLeuLysProThrHisValGluArgThrLeuPheGlyGlu 140
QY 421 GGGGATGGTTGATCGCTAGCGGTTTTCGAGACGTCGTGGTGAAGGCTGGGTGCTTATGC 480
DB 141 GlyAspGlySerSerLeuSerThrPheAspHisProLeuGlyValLeuGlyGlyLeuCys 160
QY 481 TGTGGGAGACACCTTCAACCGCTAACAAATACGCTTGTATGACAAATGAAGAGATT 540
DB 161 CysTrpGlnHisLeuGlnProLeuSerLysTyrAlaLeuTyrAlaGlnAsnGluIle 180
QY 541 CATTGTCGGCTGGCGCGGACTTTAGCTTTATCTTAATCGCGCAAAAGCCCTGGGGCCT 600
DB 181 HisPheAlaAlaTrpProSerPheSerIleTyrArgGlnAlaThrGluValLeuGlyPro 200
QY 601 GATGTCATATGATAGGCGCTCGCATCTATCGCTGAAGGGCAATGCTGCTACTAGCG 660
DB 201 GlnValAsnValAlaAlaSerArgIleTyrAlaValGluGlyGlnCysPheValLeuAla 220
QY 661 TCGTGTGCGCTCGTTTCAATCAATCGATCGATATGCTTTGTACAGATGACGAAAGCAT 720
DB 221 SerCysAlaLeuValSerProGluMetIleGluMetLeuCysThrAspGluSerLysHis 240
QY 721 GCGTTGCTTCTGGCTGGTGGTGGACATCACTGATCATAGGCGCTGATGGTGGTCACTTG 780
DB 241 SerLeuLeuGlnAlaGlyGlyGlyTyrSerArgIleIleGlyProAspGlySerAspLeu 260
QY 781 GTCGGCGCTTTCGCGAAATGAAGGGTATCTCTACGCAACCTTGATCTCGAGATA 840
DB 261 AlaArgProLeuGlyGluAsnGluGluGlyIleLeuTyrAlaThrLeuAspProAlaAla 280
QY 841 CGCATCTTGTGCTAAATGGCGGACAGACCTGCTGTCATTTATCCGTCGCCGACATTAAT 900
DB 281 ArgIleTyrAlaLysThrAlaAlaAspProAlaGlyHisIleTyrSerArgProAspValThr 300
QY 901 CGTTGCTAATATGATCGACGCCCTTAATTAACCGTAGTTGAATTGAAGTGATCTTGGCT 960
DB 301 ArgLeuLeuIleAsnArgSerAlaAsnGlnProValValGluValGlyArgGluIle--- 319
QY 961 CCTTACGCTTGGGTAAAGCCTCTGAGACGGGTGGCGCAACTGGA 1005
DB 320 -----ProAlaSerAlaGlnGlyPheGluValGlu 329

```

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - protein search, using fframe\_n2p model

Run on: April 27, 2006, 01:41:06 ; Search time 3.94745 Seconds  
(without alignments)  
2336.425 Million cell updates/sec

Title: US-09-751-299-3

Perfect score: 1826  
Sequence: 1 atcaagaagaactatcaagtc.....cgcactcgaagaatttga 1014

Scoring table:  
BLOSUM62  
Xgapop 10.0 , Xgapext 0.5  
Ygapop 10.0 , Ygapext 0.5  
Fgapop 6.0 , Fgapext 7.0  
Delop 6.0 , Delext 7.0

Searched: 232119 seqs, 45477862 residues

Total number of hits satisfying chosen parameters: 464238

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Command line parameters:

-MODEL=framed\_n2p.model -DEV=x1p  
-O=/abses/ABSSWEB.spool/US09751299/runat 26042006 090044 18835/app query.faacta\_1  
-DB=Published Applications AA New -QFMT=fastaan -SUFFIX=rapbn -MINMATCH=0.1  
-LOOPC=0 -LOOPEXT=0 -UNITS=bites -START=1 -END=-1 -MATRIX=blonum62  
-TRANS=human40.cdt -LIST=45 -DOCALIGN=200 -THR\_SCORE=pct -THR\_MAX=100  
-THR\_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=ptc -NORM=ext -HEAPSIZE=500 -MINLEN=0  
-MAXLEN=200000000 -HOST=abses02p  
-USER=US09751299 @CGN 1.1 47 @runat 26042006 090044 18835 -NCPU=6 -ICPU=3  
-NO MMAP -NEG SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG -DEV TIMEOUT=120  
-MAIN TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -Fgapop=6 -Fgapext=7  
-YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : Published Applications AA New:

1: /SIDS5/ptodata/2/pubpa/US06\_NEW\_PUB.pep:\*  
2: /SIDS5/ptodata/2/pubpa/US06\_NEW\_PUB.pep:\*  
3: /SIDS5/ptodata/2/pubpa/US07\_NEW\_PUB.pep:\*  
4: /SIDS5/ptodata/2/pubpa/US07\_NEW\_PUB.pep:\*  
5: /SIDS5/ptodata/2/pubpa/US09\_NEW\_PUB.pep:\*  
6: /SIDS5/ptodata/2/pubpa/US10\_NEW\_PUB.pep:\*  
7: /SIDS5/ptodata/2/pubpa/US11\_NEW\_PUB.pep:\*  
8: /SIDS5/ptodata/2/pubpa/US60\_NEW\_PUB.pep:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	822	45.0	356	6	US-10-537-075-7
2	780	42.7	369	6	US-10-919-182-18
3	777	42.6	369	6	US-10-919-182-16
4	776	42.5	369	6	US-10-919-182-4
5	776	42.5	369	6	US-10-919-182-6
6	774	42.4	369	6	US-10-919-182-12
7	774	42.4	369	6	US-10-919-182-14
8	765	41.9	369	6	US-10-919-182-8
9	451	24.7	333	7	US-11-096-568A-20687

10	451	24.7	351	7	US-11-096-568A-20686	Sequence 20686, A
11	446.5	24.5	330	7	US-11-096-568A-6955	Sequence 6955, Ap
12	446.5	24.5	350	7	US-11-096-568A-6954	Sequence 6954, Ap
13	323	17.7	233	7	US-11-096-568A-6956	Sequence 6956, Ap
14	274.5	15.0	193	7	US-11-096-568A-20688	Sequence 20688, A
15	126	6.9	295	7	US-11-079-463-5457	Sequence 5457, Ap
16	114.5	6.3	259	7	US-11-045-004-822	Sequence 822, Ap
17	109	6.0	296	7	US-11-045-004-2100	Sequence 2100, Ap
18	106	5.8	460	7	US-11-087-099-8682	Sequence 8682, Ap
19	103.5	5.7	493	7	US-11-096-568A-11657	Sequence 11657, A
20	103.5	5.7	643	7	US-11-096-568A-21797	Sequence 21797, A
21	100	5.5	251	7	US-11-096-568A-18174	Sequence 18174, A
22	97	5.3	388	6	US-10-497-135-19	Sequence 19, Ap
23	97	5.3	388	7	US-11-269-215-19	Sequence 19, Ap
24	97	5.3	647	7	US-11-087-099-7887	Sequence 7887, Ap
25	97	5.3	647	7	US-11-188-298-18316	Sequence 18316, A
26	96	5.3	252	7	US-11-055-822-134	Sequence 134, Ap
27	96	5.3	252	7	US-11-055-822-1120	Sequence 1120, Ap
28	95	5.2	312	7	US-11-096-568A-12515	Sequence 12515, A
29	94.5	5.2	258	7	US-11-072-512-3034	Sequence 3034, Ap
30	92	5.0	153	6	US-10-986-501-173	Sequence 173, Ap
31	92	5.0	386	6	US-10-497-135-20	Sequence 20, Ap
32	92	5.0	386	7	US-11-269-215-20	Sequence 20, Ap
33	91.5	5.0	288	7	US-11-096-568A-16873	Sequence 16873, A
34	91.5	5.0	635	7	US-11-096-568A-29628	Sequence 29628, A
35	91.5	5.0	710	7	US-11-096-568A-29627	Sequence 29627, A
36	90.5	5.0	262	7	US-11-096-568A-17194	Sequence 17194, A
37	89.5	4.9	463	7	US-11-087-099-5333	Sequence 5333, Ap
38	88.5	4.8	380	7	US-11-096-568A-27808	Sequence 27808, A
39	88.5	4.8	380	7	US-11-096-568A-21838	Sequence 21838, A
40	88.5	4.8	413	7	US-11-096-568A-20771	Sequence 20771, A
41	88	4.8	489	7	US-11-188-298-3840	Sequence 3840, Ap
42	87.5	4.8	312	7	US-11-096-568A-9235	Sequence 9235, Ap
43	86	4.7	358	7	US-11-087-099-5821	Sequence 5821, Ap
44	86	4.7	358	7	US-11-188-298-16344	Sequence 16344, A
45	85.5	4.7	293	7	US-11-096-568A-12504	Sequence 12504, A

#### ALIGNMENTS

RESULT 1  
US-10-537-075-7  
; Sequence 7, Application US/10537075  
; Publication No. US20060014291A1  
; GENERAL INFORMATION:  
; APPLICANT: Kesseler, Maria  
; APPLICANT: Zellinek, Thomas  
; TITLE OF INVENTION: L-RHAMNOSIDE-INDUCIBLE EXPRESSION SYSTEMS  
; FILE REFERENCE: 12810-00091-US  
; CURRENT APPLICATION NUMBER: US/10/537, 075  
; CURRENT FILING DATE: 2005-06-01  
; PRIOR APPLICATION NUMBER: PCT/EP2003/013367  
; PRIOR FILING DATE: 2003-11-27  
; PRIOR APPLICATION NUMBER: DE 102 56 381.0  
; PRIOR FILING DATE: 2002-12-02  
; NUMBER OF SEQ ID NOS: 19  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 7  
; LENGTH: 356  
; TYPE: PRT  
; ORGANISM: Alcaligenes faecalis  
US-10-537-075-7

Alignment Scores:  
Pred. No.: 2,43e-76 Length: 356  
Score: 822.00 Matches: 160  
Percent Similarity: 65.1% Conservative: 56  
Best Local Similarity: 48.2% Mismatches: 114  
Query Match: 45.0% Indels: 2  
DB: Gaps: 1  
US-09-751-299-3 (1-1014) x US-10-537-075-7 (1-356)

```
Qy 4 AAAGAGCTATCAAGTCGCTGGTGCAAGCGCCCGATCTACATGATTGGAGGCG 63
Db 4 ArglyslleValarglaalalaValaGlnalalaserProdanrYrAspLeuAlaThr 23
Qy 64 ACGGTGACAAACCATTTAGATTGATGAGAGACGACGCTAAATGCTCGTGTATC 123
Db 24 GlyValAspYsThrIleGluLeuAlaArgGlnAlaArgAspGluGlyCyAspLeuIle 43
Qy 124 GCCTTCCGAAACTTGGATTCCAGGCTACCCAGGTTCCTTGGCTTGAATCAGCA 183
Db 44 ValPheGlyGluThrLeuProGlyYrProPheIstValITrPheGlyAlaProAla 63
Qy 184 TGGCAATGCAATTTGACGCGCAATACAGATAGAGATGATGAGTGGAGCCCTCA 243
Db 64 TrpSerLeuYrYrSerAlaArgYrYrAlaAsnSerLeuSerLeuAspSerAlaGlu 83
Qy 244 GCTAAGCGCATTTGATGATGACGCGCAAGCGGTTGGGAATCATGTCACCTTGGGATGAGT 303
Db 84 PheGlnArgIleAlaGlnAlaAlaArgThrLeuGlyIlePheIleAlaLeuGlyYrSer 103
Qy 304 GAACGGTCGGTGCGACCTTTACATCAGTCAGTGGTTCATAGGCGATATGGTGACAC 363
Db 104 GluArgSerGlyGlySerLeuYrLeuGlyGlnCyAspLeuIleAspAspGlyGlyGluMet 123
Qy 364 ATTGGGGCGCGGCAAGTTGAAACCTTCTTTGTGAAAGTACCTTGTGCGGCAAGG 423
Db 124 LeuTrpSerAlaGlyGlyLeuYrProThrIstValaGlnaGlnThrValPheGlyGluGly 143
Qy 424 GATGGTTTCATGCTAAGCGGTTTTCGAGAGCTCTGTGGAAGCGTGGGCTTATGCTGT 483
Db 144 TyrAlaArgAspLeuIleValaSerAspThrGluLeuGlyArgValaGlyAlaLeuCySer 163
Qy 484 TGGAGACCTTCAACGCTAAACAAATACGCTTGTATGACAAATAGAGAGATTCAT 543
Db 164 TrpGluIstLeuSerProLeuSerYrYrAlaLeuYrSerGlnIstValaIleIst 183
Qy 544 TGTGCGGCTTGGCGGAGCTTTAGCCTTATCTTAATGCGGCAAGCCCTGAGGCTGAT 603
Db 184 IleAlaIstAlaTrpProSerPheSerLeuYrSerGluGlnAlaIstAlaLeuSerAlaYs 203
Qy 604 GTCAATGAGCGGCTCTGCAATCTATACGCTTGAAGGCAATGCTTGTATGCGTTCG 663
Db 204 ValAsnMetAlaAlaSerGlnIleYrSerValaGluGlyGlnCyPheThrIleAlaIa 223
Qy 664 TGTGGGCTGTTTCAATCATGATGATGATGCTTGTATGATGACGAAGATGATGCG 723
Db 224 SerSerValaIstGlnGluThrLeuAspMetLeuGluValaGlyGlnIstAsnAlaPro 243
Qy 724 TTGCTTCTGGCTGTGTGACACTCAAGTATCATAGGCGCTGATGCTGTGATCTGTG 783
Db 244 LeuLeuYsValaGlyGlySerSerMetIlePheAlaProAspGlyArgThrLeuAla 263
Qy 784 GCGGCTTGGCGGAAATGAAGAGGATTTCTTACGCAAACTTGAATCTTGAAGTACG 843
Db 264 ProYrLeuProIstAspAlaGluGlyLeuIleIleAlaAspLeuAsnMetGluGluIle 283
Qy 844 ATTCCTTGTAAATGCGGAGACCTTGTGCTATTTGCGCGCGCAATTAATCTGCG 903
Db 284 AlaPheAlaYsAlaIleAsnAspProValaGlyIstYrSerYsProGlnAlaThrArg 303
Qy 904 TTGCTAATAGATCGGAGCGCTAAATTAACCGTATGTTGAATGTAAGGATCTTCGCTCT 963
Db 304 LeuValLeuAspLeuGlyIstAspAspProMetThrArgValIstSer-----LysSer 321
Qy 964 TACGCTTTGGTAAAGCGTCTGACAGCGGCTGCGCA 999
Db 322 ValThrArgGluGluAlaProGluGlnIstValaGln 333
```

RESULT 2  
US-10-919-182-18  
Sequence 18, Application US/10919182  
Publication No. US2006003532A1

```
GENERAL INFORMATION:  
APPLICANT: E.I. duPont de Nemours and Company, Inc.  
APPLICANT: Di Cosimo, Robert  
APPLICANT: Payne, Mark  
APPLICANT: O'Keefe, Daniel  
TITLE OF INVENTION: PRODUCTION OF 3-HYDROXYCARBOXYLIC ACID USING NITRILASE MUTANTS  
FILE REFERENCE: CL2584 US NA  
CURRENT APPLICATION NUMBER: US/10/919,182  
CURRENT FILING DATE: 2004-08-16  
NUMBER OF SEQ ID NOS: 18  
SOFTWARE: PatentIn version 3.2  
SEQ ID NO 18  
LENGTH: 369  
TYPE: PRT  
ORGANISM: Artificial sequence  
FEATURES:  
OTHER INFORMATION: Mutant nitrilase with Phe168 to Leu change  
US-10-919-182-18  
Alignment Scores:  
Pred. No.: 5,16e-72 Length: 369  
Score: 780.00 Matches: 151  
Percent Similarity: 67.3% Conservative: 61  
Best Local Similarity: 47.9% Mismatches: 99  
Query Match: 42.7% Indels: 4  
DB: Gaps: 3  
US-09-751-299-3 (1-1014) x US-10-919-182-18 (1-369)  
Qy 22 GCCTGCTGCAAGCGCCCGCATCTACATGATTGGAGCGGACGATGCAAAACAT 81  
Db 11 AlathrValaGlnAlaGluProValaITrPheuAspAlaAspAlathrIleAspYrSerIle 30  
Qy 82 GAGTGTAGAGAGACGACGCAATTAATGCTGCTGCTGCTGCTTCCGAAACTGG 141  
Db 31 GlyIleIleGluGlnAlaAlaGlnIstYsGlyAlaSerLeuIleAlaPheProGluValaPhe 50  
Qy 142 ATTCAGGCTTACCCATGCTTCTTGGCTTGAATGCTGCTGCTGCTGCTGCTGCTGCTG 201  
Db 51 IleProGlyYrProYrYrTrpAlaITrPheGlyYAspValYsYrSerLeuSerPheThr 70  
Qy 202 CGCCAAATACATGAGAACTCATGTTGAGATTGAGTGGCCCTCAAGCTAAGCCATTTCA 261  
Db 71 SerYrYrIstGlnIstAsnSerLeuGluLeuGlyYAspAspArgMetAlaGluLeu 90  
Qy 262 GCAAGCGGCTTGGGAATCATGATGCTACCCCTGGGAGATGATGAACCGGCTCGTGACCC 321  
Db 91 AlaIstArgArgAsnYsIleAlaLeuValMetClyYrSerGlnValaGluAlaGlySer 110  
Qy 322 CTTTACATCACTAGTGGTTTCATAGGCGATATGTTGACACCATTTGGGCGCGGAAG 381  
Db 111 ArgYrLeuSerGlnValaPheIleAspGluArgGlyGluIleValaIstAsnArgArgYs 130  
Qy 382 TTGAACCTTACTTTTGTGAAACGTAATCTTGTGGCGGCAAGGGATGTTTATGCTAGCG 441  
Db 131 LeuYsProThrIstValaGluArgThrIleYrGlyGluGlyAsnGlyThrAspPheLeu 150  
Qy 442 GTTTTGAAGCGTGTGTTGAAGCGTGGGCTTATGCTGTTGGGAGCACTTCAACCG 501  
Db 151 ThrIstAspPheAlaPheGlyArgValaGlyIleuAsnYsTrpGluIstLeuGlnPro 170  
Qy 502 CTAAACAAATACGCTTTGTATGCAAAATGAAGAGATTCATTTGGCGCTTGGCGGAC 561  
Db 171 LeuSerYsPheMetClyYrSerLeuGlyGluGlnValaIstValaIstSerTrpProAla 190  
Qy 562 TTTTACG--CTTATCTTAATGCGGCGAAAGCCTTGGGCTGATGTTCATGTAGCGGCC 618  
Db 191 MetSerProLeuGlnProAspValaPheGln--LeuSerIleGlnAlaAsnAlaThrVal 209  
Qy 619 TCTGGAATCATGCGGCTTGAAGGCAATGCTTCTGATGAGCGTGTGCGCTGTTTCA 678  
Db 210 ThrArgSerYrAlaIleGluGlyGlnThrPheValaLeuCySerThrGlnValaIleGly 229
```



```
US-09-751-299-3 (1-1014) x US-10-919-182-4 (1-369)
QY 22 GCCTGCGTGAAGCCGCCGATCTACATGATTTGGAGGCGACGCGTGAACAACCAT 81
Db 11 AlathrValGlnAlaGlnProValTrrPleuAspAlaPheValThrIleAspLysSerIle 30
QY 82 GAGTGTATGTAAGAAGACGACGCTAATAATAGCTGCTGATGAGCCCTTCCGGAACCTTGG 141
Db 31 GlyIleIleGlnGlnAlaAlaGlnLysGlyAlaSerLeuIleAlaPheProGlnValPhe 50
QY 142 ATTCAGGCTACCCATGCTTCTTGGCTTACTCACCAGCATGGGCAATGCAATTTGTA 201
Db 51 IleProGlyTrrProTrrTrrPalaTrrPleuGlyAspValLysTrrSerLeuSerPheThr 70
QY 202 CGCCATACCATGTAAGAACTCATTTGAGATGGCCCTCAAGGCTAAGCCGATTTTCAGAT 261
Db 71 SerArgTrrHisGlnAsnSerLeuGlnLysAlaAspArgMetArgArgLeuGlnLeu 90
QY 262 GCACGCAAGCGGTTGGGAATGATGTCACCCCTGGGGAGATGTAACGCGTGGCGACCC 321
Db 91 AlaAlaArgArgAsnLysIleAlaLeuValMetGlyTrrSerGlnArgGlnAlaGlySer 110
QY 322 CTTCATCATGCTCATGCTTCATAGCGCATATGATGTCACACCATTTGGGCGCGGCAAG 381
Db 111 ArgTrrLeuSerGlnValPheIleAspGlnArgGlyGlnIleValAlaAsnArgArgLys 130
QY 382 TTGAAACCTACTTTTGTGGAACGTACTTTGTCGCGCAAGGGGAGATGGTTCATCGCTACG 441
Db 131 LeuLysProThrHisValGlnArgThrIleTrrGlyGlnGlyAsnGlyThrAspPheLeu 150
QY 442 GTTTTCGACGTCGTGTGGAAGCGTGGGCTTATGCTGTGGAGACACCTTCAACCG 501
Db 151 ThrHisAspPheAlaPheGlyArgValGlyGlyLeuAsnCystrPglunHisPheGlnPro 170
QY 502 CTAACAAAATACGCTTGTGATGACCAAAATGAAAGATTCATTTGCGCGCTTGGCCGAGC 561
Db 171 LeuSerLysPheMetMetLysSerLeuGlyGlnGlnValHisValAlaSerTrrProAla 190
QY 562 TTTAGC---CTTTATCCTTAATGCGCGCAAGCCCTGGGGCGCTGATGCAATGTAGCGGCC 618
Db 191 MetSerProLeuGlnProAspValPheGln--LeuSerIleGlnAlaAsnAlaThrVal 209
QY 619 TCTGGAATCTATGCGCTGTAAGGCGCAATGCTTGTACTAGACGTCGTGCGCTGTTTCA 678
Db 210 ThrArgSerTrrAlaIleGlnGlyGlnThrPheValLeuCyserThrGlnValIleGly 229
QY 679 CAATTCATGATCATATGCTTTGTACAGATGACGAAAAGCATGCGTTCCTTGGCTGGT 738
Db 230 ProSerAlaIleGlnThrPheCysLeuAsnAspGlnGlnArgAlaLeuLeuProGlnGly 249
QY 739 GGTGGAACATCATGATGATAGGGCTGATGCTGGTGGTGGTGGCGGCTTGGCGGAA 798
Db 250 CysGlyTrrPalaArgIleTrrGlyProAspLysSerGlnLeuValLysProLeuAlaGln 269
QY 799 AATGAAGAGGTATTTCTTACGCAAACTTGTAGCTTGGAGTACGACCTTGTAAATG 858
Db 270 AspAlaGlnGlyIleLeuTrrAlaGlnIleAspLeuGlnGlnIleLeuAlaLysAla 289
QY 859 GCGGCAAGACCTGCTGCTATTTATCCGCTCCGACATTAATCGCTTGTCTAATGATGCC 918
Db 290 GlnAlaAspProValGlyHisLysTrrSerArgProAspValLeuSerValGlnPheAspPro 309
QY 919 AGCCCTAAATTAACGGTA-----GTGAAATTTGAAGTGAATCTT 957
Db 310 ArgAsnHisTrrProValHisArgIleGlyIleAspGlyArgLeu 324

RESULT 5
US-10-919-182-6
; Sequence 6, Application US/10919182
; Publication No. US2006003532A1
; GENERAL INFORMATION:
; APPLICANT: E.I. duPont de Nemours and Company, Inc.
; APPLICANT: DI Cosimo, Robert

; APPLICANT: Payne, Mark
; APPLICANT: O'Keefe, Daniel
; TITLE OF INVENTION: PRODUCTION OF 3-HYDROXYCARBOXYLIC ACID USING NITRILASE MUTANTS
; FILE REFERENCE: CL2584 US NA
; CURRENT APPLICATION NUMBER: US/10/919,182
; CURRENT FILING DATE: 2004-08-16
; NUMBER OF SEQ. ID NOS: 18
; SOFTWARE: PatentIn version 3.2
; SEQ. ID NO: 6
; LENGTH: 369
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Mutant nitrilase B2 and H9
US-10-919-182-6

Alignment Scores:
Pred. No.: 1,33e-71 Length: 369
Score: 776.00 Matches: 150
Percent Similarity: 67.0% Conservative: 61
Best Local Similarity: 47.6% Mismatches: 100
Query Match: 42.5% Indels: 4
Gaps: 3

US-09-751-299-3 (1-1014) x US-10-919-182-6 (1-369)
QY 22 GCCTGCGTGAAGCCGCCGATCTACATGATTTGGAGGCGACGCGTGAACAACCAT 81
Db 11 AlathrValGlnAlaGlnProValTrrPleuAspAlaPheValThrIleAspLysSerIle 30
QY 82 GAGTGTATGTAAGAAGACGACGCTAATAATAGCTGCTGATGAGCCCTTCCGGAACCTTGG 141
Db 31 GlyIleIleGlnGlnAlaAlaGlnLysGlyAlaSerLeuIleAlaPheProGlnValPhe 50
QY 142 ATTCAGGCTACCCATGCTTCTTGGCTTACTCACCAGCATGGGCAATGCAATTTGTA 201
Db 51 IleProGlyTrrProTrrTrrPalaTrrPleuGlyAspValLysTrrSerLeuSerPheThr 70
QY 202 CGCCATACCATGTAAGAACTCATTTGAGATGGCCCTCAAGGCTAAGCCGATTTTCAGAT 261
Db 71 SerArgTrrHisGlnAsnSerLeuGlnLysAlaAspArgMetArgArgLeuGlnLeu 90
QY 262 GCACGCAAGCGGTTGGGAATGATGTCACCCCTGGGGAGATGTAACGCGTGGCGACCC 321
Db 91 AlaAlaArgArgAsnLysIleAlaLeuValMetGlyTrrSerGlnArgGlnAlaGlySer 110
QY 322 CTTCATCATGCTCATGCTTCATAGCGCATATGATGTCACACCATTTGGGCGCGGCAAG 381
Db 111 ArgTrrLeuSerGlnValPheIleAspGlnArgGlyGlnIleValAlaAsnArgArgLys 130
QY 382 TTGAAACCTACTTTTGTGGAACGTACTTTGTCGCGCAAGGGGAGATGGTTCATCGCTACG 441
Db 131 LeuLysProThrHisValGlnArgThrIleTrrGlyGlnGlyAsnGlyThrAspPheLeu 150
QY 442 GTTTTCGACGTCGTGTGGAAGCGTGGGCTTATGCTGTGGAGACACCTTCAACCG 501
Db 151 ThrHisAspPheAlaPheGlyArgValGlyGlyLeuAsnCystrPglunHisPheGlnPro 170
QY 502 CTAACAAAATACGCTTGTGATGACCAAAATGAAAGATTCATTTGCGCGCTTGGCCGAGC 561
Db 171 LeuSerLysPheMetMetLysSerLeuGlyGlnGlnValHisValAlaSerTrrProAla 190
QY 562 TTTAGC---CTTTATCCTTAATGCGCGCAAGCCCTGGGGCGCTGATGCAATGTAGCGGCC 618
Db 191 MetSerProLeuGlnProAspValPheGln--LeuSerIleGlnAlaAsnAlaThrVal 209
QY 619 TCTGGAATCTATGCGCTGTAAGGCGCAATGCTTGTACTAGACGTCGTGCGCTGTTTCA 678
Db 210 ThrArgSerTrrAlaIleGlnGlyGlnThrPheValLeuCyserThrGlnValIleGly 229
QY 679 CAATTCATGATCATATGCTTTGTACAGATGACGAAAAGCATGCGTTCCTTGGCTGGT 738
Db 230 ProSerAlaIleGlnThrPheCysLeuAsnAspGlnGlnArgAlaLeuLeuProGlnGly 249
```

QY 382 TTGAAACCACTTTGTTGAAACGATCTTTGCGCGAAGGAGAGGTTGATCGCTACGC 441  
 Db 131 LeuLysProIlnHlsValGIuIdrItnrIleYrIolYgluYlYnAsnGIYlnHAspPheLeu 150  
 QY 442 GTTTTCGAGACGCTCTGTGGAAGGCGTGCGCTTATGCTGTGGAGCACTTCAACG 501  
 Db 151 ThrlsAspPheAlaPheGIYArGValGIYlYleuYnAsnCystrGIuHlsPheGlnPro 170  
 QY 502 CTAAACAAATATACGCTTTGTATGCACAAATATGAAGATTCTTTGGCGGCTGGCCGAGC 561  
 Db 171 LeuSerLysPheMetMetLysrIeueGIYgluGlnAlaHlsValAlAserTrpProAla 190  
 QY 562 TTTTAC---CTTTATCTTAATCGCGCGCAACCCCGGGGCGCTGATGTCATGTAGCGGCC 618  
 Db 191 MetSerProIleuGlnProAspValPheGln--LeuSerIleGIuAlaAsnAlaThrVal 209  
 QY 619 TCTCGAATCTATGCCGTGGAAGGCGCAATGCTTCGTACTAGCGCTGTGCGCTCGTTTCA 678  
 Db 210 CysAtGserTrYrAlaIleGluGIYgluIlnrPheValIleuCysserThrGlnValIleGIY 229  
 QY 679 CAATCCATGATGCATATGCTTTGTACAGATACGAAAGCATGCGTTCCTTCGCTGCT 738  
 Db 230 ProSerAlaIleGIuIlnrPheCysLeuAsnAspGIuIlnrAlaIleuLeuProGlnGIY 249  
 QY 739 GGTGACACCTACGATATGAGGCGCTGATGCTGTGATGCTTCGTGCGCGCTTCGCCGA 798  
 Db 250 CysGIYTrpAlaAspIleTrYrGIYProAspGIYserGIuIleuAlaLysProIleuAlaGIu 269  
 QY 799 AATGAAGAGGATATCTCTACGCAACCTTGATCTGAGATGACGATCTTGCTAAATG 858  
 Db 270 AspAlaGIuGIYlIleuTrYrAlaGIuIlnrLeuAspGIuGIuIlnrLeuAlaLysAla 289  
 QY 859 GCGGAGACCTGCTGGCTGATTTATTCGCTGCCGACATYACTGCGCTTGCTAATGATGCG 918  
 Db 290 GIYAlaAspProValGIYHlsTrYserTrAspGIYProAspValIleuSerValGlnPheAspPro 309  
 QY 919 AGCCCTAATTAACCGGTA-----GTGGAATTGAAGGTATGCTT 957  
 Db 310 ArgAsnHlsThrProValHlsArgIleGIYlIleAspAlaArgLeu 324  
 RESULT 7  
 US-10-919-182-14  
 ; Sequence 14, Application US/10919182  
 ; Publication No. US20060035352A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: E.I. duPont de Nemours and Company, Inc.  
 ; APPLICANT: Di Cosimo, Robert  
 ; APPLICANT: Payne, Mark  
 ; APPLICANT: O'Keefe, Daniel  
 ; TITLE OF INVENTION: PRODUCTION OF 3-HYDROXYCARBOXYLIC ACID USING NITRILASE MUTANTS  
 ; FILE REFERENCE: CL2584 US NA  
 ; CURRENT APPLICATION NUMBER: US/10/919,182  
 ; CURRENT FILING DATE: 2004-08-16  
 ; NUMBER OF SEQ ID NOS: 18  
 ; SOFTWARE: PatentIn version 3.2  
 ; SEQ ID NO 14  
 ; LENGTH: 369  
 ; TYPE: PRT  
 ; ORGANISM: Artificial sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: Mutant nitrilase with Phe168 to Lys change  
 US-10-919-182-14  
 Alignment Scores:  
 Pired. No.: 2,14e-71 Length: 369  
 Score: 774.00 Matches: 150  
 Percent Similarity: 67.0% Conservative: 61  
 Best Local Similarity: 47.6% Mismatches: 100  
 Query Match: 42.4% Indels: 4  
 DB: 6 Gaps: 3

```
QY 22 GCTGCGTGAAGCCGCCGATCTACATGATTTGGAGCGGACGAGCAAAACATT 81
Db 11 AlathrValGlnAlaGluProValTrpLeuAspAlaThrIleAspLysSerIle 30
QY 82 GAGTTGATGAAGAAGACAGCAATGATATGCTGCTGATCGCTTTCCGAAACTGG 141
Db 31 GlyIleIleGluGlnAlaIleGlnLysIleAspSerIleAlaPheProGluValPhe 50
QY 142 ATTCCAGGCTACCCATGGTCTTTGCTTGGCTTGCATCCACAGCATGGGCAATGTTGTA 201
Db 51 IleProGlyTrpProGlyTrpAlaTrpLeuGlyAspValLysTrpSerLeuSerPheThr 70
QY 202 CGCCAAATACCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 261
Db 71 SerArgTrpHisGlnAspSerLeuGlnLysAspAspArgMetArgArgLeuGlnLeu 90
QY 262 GCAGCCAAAGCGGTTGGAGATCATGATGATGATGATGATGATGATGATGATGATGATGAT 321
Db 91 AlaIleArgAspGlnLysIleIleAlaLeuValMetGlyTrpSerGluArgGluAlaGlySer 110
QY 322 CTTTACATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 381
Db 111 ArgTrpLeuSerGlnValPheIleAspGlnArgGlyGluIleValAlaAsnArgTrpGlys 130
QY 382 TTGAACCTACTTTGTTGAACGTAATGTTGTTGCGCGAAGGAGATGTTCAATCGCTACG 441
Db 131 LeuArgProThrHisValGlnArgThrIleTrpGlyGlnGlyAsnGlyThrAspPheLeu 150
QY 442 GTTTTCGAGAGCTGTGTTGGAAGGCTGGGCTTATGCTGTGTTGGAGACCTTTAAACG 501
Db 151 ThrHisAspPheAlaPheGlyArgValGlyGlyLeuAsnCysTrpGlnHisLysGlnPro 170
QY 502 CTAACAAAATACGCTTTGATGACAAATGATGATGATGATGATGATGATGATGATGATGAT 561
Db 171 LeuSerLysPheMetMetLysSerLeuGlnGlnValHisValAlaSerTrpProAla 190
QY 562 TTTTAC---CTTTATCTTAATGCGGCGAAGCCCTGGGCGCTGAGTCAATGATGACGGC 618
Db 191 MetSerProLeuGlnProAspValPheGln--LeuSerIleGlnAlaAsnAlaThrVal 209
QY 619 TCTGCAATCTATGCGGTTGAAGGAGCATGCTGCTGATGATGATGATGATGATGATGATGAT 678
Db 210 ThrArgSerTrpAlaIleGlnGlyGlnThrPheValLeuCysSerThrGlnValIleGly 229
QY 679 CAATCCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 738
Db 230 ProSerAlaIleGluThrPheCysLeuAsnAspGlnArgAlaLeuLeuProGlnGly 249
QY 739 GGTGACACATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 798
Db 250 CysGlyTrpAlaArgIleTrpGlyProAspGlySerGlnLeuAlaLysProLeuAlaGln 269
QY 799 AATGAAGAGGATATTCTTACGCAAACTTGAATCCTGATGATGATGATGATGATGATGATGAT 858
Db 270 AspAlaGlnGlyIleLeuTrpAlaGlnIleAspLeuGlnGlnIleLeuLeuAlaLysAla 289
QY 859 GGGGAGAGCCCTGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 918
Db 290 GlyAlaAspProValGlyHisIleTrpSerArgProAspValLeuSerValGlnPheAspPro 309
QY 919 AGCCCTTAATATTCGCGTA-----GTTGAATTTGAAGATGATCTT 957
Db 310 ArgAsnHisThrProValHisArgGlyIleAspGlyArgLeu 324
```

```
APPLICANT: O'Keefe, Daniel
TITLE OF INVENTION: PRODUCTION OF 3-HYDROXYCARBOXYLIC ACID USING NITRILASE MUTANTS
FILE REFERENCE: C12584 US NA
CURRENT APPLICATION NUMBER: US/10/919,182
CURRENT FILING DATE: 2004-08-16
NUMBER OF SEQ ID NOS: 18
SOFTWARE: PatentIn version 3.2
SEQ ID NO: 8
LENGTH: 369
TYPE: PRT
ORGANISM: Artificial sequence
FEATURE:
OTHER INFORMATION: Mutant nitrilase B4
US-10-919-182-8

Alignment Scores:
Pred. No.: 1,81e-70 Length: 369
Score: 765.00 Matches: 150
Percent Similarity: 66.0% Conservative: 58
Best Local Similarity: 47.6% Mismatches: 103
Query Match: 41.9% Indels: 4
DB: 6 Gaps: 3

US-09-751-299-3 (1-1014) x US-10-919-182-8 (1-369)

QY 22 GCTGCGTGAAGCCGCCGATCTACATGATTTGGAGCGGACGAGCAAAACATT 81
Db 11 AlathrValGlnAlaGluProValTrpLeuAspAlaThrIleAspLysSerIle 30
QY 82 GAGTTGATGAAGAAGACAGCAATGATATGCTGCTGATCGCTTTCCGAAACTGG 141
Db 31 GlyIleIleGluGlnAlaIleGlnLysIleAspSerIleAlaPheProGluValPhe 50
QY 142 ATTCCAGGCTACCCATGGTCTTTGCTTGGCTTGCATCCACAGCATGGGCAATGTTGTA 201
Db 51 IleProGlyTrpProGlyTrpAlaTrpLeuGlyAspValLysTrpSerLeuSerPheThr 70
QY 202 CGCCAAATACCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 261
Db 71 SerArgTrpHisGlnAspSerLeuGlnLysAspAspArgMetArgArgLeuGlnLeu 90
QY 262 GCAGCCAAAGCGGTTGGAGATCATGATGATGATGATGATGATGATGATGATGATGATGAT 321
Db 91 AlaIleArgAspGlnLysIleIleAlaLeuValMetGlyTrpSerGluArgGluAlaGlySer 110
QY 322 CTTTACATCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 381
Db 111 ArgTrpLeuSerGlnValPheIleAspGlnArgGlyGluIleValAlaAsnArgTrpGlys 130
QY 382 TTGAACCTACTTTGTTGAACGTAATGTTGTTGCGCGAAGGAGATGTTCAATCGCTACG 441
Db 131 LeuArgProThrHisValGlnArgThrIleTrpGlyGlnGlyAsnGlyThrAspPheLeu 150
QY 442 GTTTTCGAGAGCTGTGTTGGAAGGCTGGGCTTATGCTGTGTTGGAGACCTTTAAACG 501
Db 151 ThrHisAspPheAlaPheGlyArgValGlyGlyLeuAsnCysTrpGlnHisLysGlnPro 170
QY 502 CTAACAAAATACGCTTTGATGACAAATGATGATGATGATGATGATGATGATGATGATGAT 561
Db 171 LeuSerLysIleMetMetLysSerLeuGlnGlnValHisValAlaSerTrpProAla 190
QY 562 TTTTAC---CTTTATCTTAATGCGGCGAAGCCCTGGGCGCTGAGTCAATGATGACGGC 618
Db 191 MetSerProLeuGlnProAspValPheGln--LeuSerIleGlnAlaAsnAlaThrVal 209
QY 619 TCTGCAATCTATGCGGTTGAAGGAGCATGCTGCTGATGATGATGATGATGATGATGATGAT 678
Db 210 ThrArgSerTrpAlaIleGlnGlyGlnThrPheValLeuCysSerThrGlnValIleGly 229
QY 679 CAATCCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 738
Db 230 ProSerAlaIleGluThrPheCysLeuAsnAspGlnArgAlaLeuLeuProGlnGly 249
```





US-09-751-299-3 (1-1014) x US-11-096-568A-20686 (1-351)

```

QY 16 AAGTGGCCGCGGCAAGCGCCGCGATCATGATGATTTGGAGCGGCGAAGAA 75
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
   31 ArgValThrValValGlnIleSerSerValPheTyrAspThrProIleThrLeuAspLys 50
QY 76 ACATGAGTTGATGAAAGAGCAGCAGCTAATATGCTGCTGATCCGCTTCCGAA 135
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
   51 AlaGlnLysLeuValAlaGlnIleAlaGlyTyrGlySerGlnLeuValLeuPheProGln 70
QY 136 ACTTGAAATCCAGGCTACCA-----TGATTCTTTGGCTTGACTACCA 180
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
   71 ValPheValGlyGlyTyrProIleGlySerThrPheGlyLeuValValGlyAsnArgThr 90
QY 181 GCATGGGCAATGCAATTTGTAAGCCCAATACATAGAAATCATTTGAGTTGATGGCCCT 240
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
   91 AlaLysGlyLysGlnAspPheGlnLysTyrIleSerIleAlaIleAspValProGlyPro 110
QY 241 CAAGCTAAGCGCATTTGATGATGCAAGCGGTTGGAAATCATGATCATCCCTGGGATG 300
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
   111 GlnValSerArgLysSerAlaLeuAlaGlyLysTyrLysValPheLeuValIleGlyVal 130
QY 301 AGTGAACGGGTGGTGGACACCTTTATCATCATGATGCTTCAATAGCGGATATGTTGAC 360
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
   131 ValGlnArgAlaGlyTyrThrLeuTyrAsnThrValLeuSerPheAspProLeuGlyLys 150
QY 361 ACCATTGGGGCGCGGCAAGTTGAACCTACTTTGTTGAAACGTACTTTGTCGGCGAA 420
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
   151 TyrLeuGlyLysHisArgLysValMetProThrAlaLeuGlnArgValPheTrpGlyPhe 170
QY 421 GGGAGATGTTATCGCTAGCGGTTTTCAGACGCTGCTTTGGAAGCGCTGGGCTTATGC 480
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
   171 GlyAspGlySerThrIleProValTyrAspThrProIleGlyLysMetGlyAlaLeuIle 190
QY 481 TGTGGAGACACCTTCAACCGCTTAACAAATAGCGCTTGTATGACAAATAGAGAGAT 540
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
   191 CysTrpGlnAsnArgMetProLeuLeuAlaGlnThrIleMetTyrAlaLysGlyIleGlnIle 210
QY 541 CATTTGCGGCTTGGCGGAGCTTTAGCTTTATCTTAATGCGGCAAGCCCTGGGGCCT 600
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
   211 TyrCysAla-----ProThrValAspCysMetProThrTrp----- 222
QY 601 GATGTCAATGTACGCGCTCGAATCTATACCGTTGAAGGCAATGCTTGGTACTAGCG 660
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
   223 -----LeuSerSerMetThrHisIleAlaLeuGlnGlyCysPheValLeuSer 239
QY 661 TCGGTGCGCTCGTTTCAATCCATCATGATGAT-----ATGCTTTGTACA 705
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
   240 AlaCysGlnPheCysArgLysAsnTyrProProProProGlnTyrThrPheCysGly 259
QY 706 GATGACGAAAG-----CATGCGTCTTCTGCTGATGATGACACTCAGCTATTC 756
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
   260 LeuGlnGlnGlnProSerProGlnSerValValCysSerGlyGly-----SerValIle 277
QY 757 ATAGGCGCTGATGTTGATGCTTGTGCGCGCTTTCGCAAAATGAAGAGGATATTTC 816
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
   278 IleSerProLeuGlyThrValLeuAlaGlyProAsnTyrGlnSer---GlnAlaLeuLeu 296
QY 817 TACGCAAACTTGATCTGAGTAGAGTACGATCTTGTCTAAATGGCGAGACCTGTGCT 876
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
   297 ThrIleAspLeuAspLeuGlyGlnIleValAlaArgAlaLysPheAspPheValValGly 316
QY 877 CATTAATCCGCTCCGACATTAATCGCTTGAATAGATCGACCCCTAAATTAACGGTA 936
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
   317 HisTyrSerArgProGlnValLeuSerLeuValLysSerAspProLys---ProAla 335
QY 937 GTT 939
   |||
   336 Val 336

```

RESULT 11  
US-11-096-568A-6955  
/ Sequence 6955, Application US/11096568A

```

/ Publication No. US20060048240A1
/ GENERAL INFORMATION:
/ APPLICANT: Alexandrov, Nikolai et al.
/ TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
/ TITLE OF INVENTION: Theby
/ FIDR REFERENCE: 2750-1592PUS2
/ CURRENT APPLICATION NUMBER: US/11/096,568A
/ CURRENT FILING DATE: 2005-04-01
/ NUMBER OF SEQ ID NOS: 34471
/ SEQ ID NO 6955
/ LENGTH: 330
/ TYPE: PRT
/ ORGANISM: Glycine max
/ FEATURE:
/ NAME/KEY: misc_feature
/ LOCATION: (1)..(330)
/ OTHER INFORMATION: Ceres Seq. ID no. 15169318
/ US-11-096-568A-6955

Alignment Scores:
Pred. No.: 1,096-37 Length: 330
Score: 446.50 Matches: 111
Percent Similarity: 51.7% Conservative: 60
Best Local Similarity: 33.5% Mismatches: 113
Query Match: 24.5% Indels: 47
DB: Gaps: 10

US-09-751-299-3 (1-1014) x US-11-096-568A-6955 (1-330)

QY 13 ATCAAGTGGCCGCTGCGTGAAGCGCCGCGATCATGATGATTTGGAGCGGCGTGC 72
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
   10 ValArgAlaThrValValGlnIleSerSerThrIlePheTyrAspThrProIleThrLeuAsp 29
QY 73 AAAACATTGATGATGATGAAAGAGCAGCAGCTAATATGCTGCTGATCCGCTTCCG 132
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
   30 LysAlaGlnArgLeuValLeuAlaGlnIleThrSerTyrGlySerGlnLeuValAlaPhePro 49
QY 133 GAAACTTGATTCAGGCTTACCATGCTTTCTTGGCTTGATGATGACGAGGCAATG 192
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
   50 GlnAlaPheValGlyGlyTyrPro-----ArgGlySerIlePheGlyLeu 64
QY 193 CAATTT-----GTACGCAATACCTAGAACTCA 222
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
   65 SerIleGlyAsnArgThrValLysGlyArgGlnGlnPheArgLysTyrHisSerAlaAla 84
QY 223 TTGAGTTGATGAGCCCTCAAGCTAAGCGCATTTCAATGATGACGCAAGCGGTTGGATC 282
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
   85 IleAspValProGlyProGlnValAspArgLeuAlaIleMetAlaGlyTyrLysVal 104
QY 283 ATGGTACCCGCGGATGATGAAAGCGGTGCGGACCTTTATCATCACTAGCTGTTTC 342
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
   105 HisLeuValMetGlyValIleGlnArgAspGlyTyrThrLeuTyrCysThrValLeuPhe 124
QY 343 ATAGCGATTAATGTGACACCATTTGGGGCCCGGCAAGTTGAACCTACTTTGTTGAA 402
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
   125 PheAspSerGlnGlnHisTyrLeuGlyLysHisArgLysIleMetProThrAlaLeuGln 144
QY 403 CGTACTTTGTCGCGGAGGAGATGTTCAATGCTAGCGGTTTTCGAGAGCTGTTGGA 462
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
   145 ArgValIleTrpGlyPheGlyAspGlySerThrIleProValPheGlnThrProValGly 164
QY 463 AGCTGGGTGGCTTATGCTGTTGGAGACACTTCAACCGCTTAACAAATAGAGCTTTGAT 522
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
   165 LysIleGlyAlaAlaIleCysTrpIleAsnArgMetProLeuLeuValThrIleMetTyr 184
QY 523 GCACAAATAGAGATTCATGATGCGGCTTGGCGGACCTTGAACCTTATCTTAATCG 582
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
   185 AlaLysGlyValGlnLeuTyrCysAla-----ProThrAla 196
QY 583 GCGAAAGCCCTGGGCGCTGATGATGATGATGCGGCTTCAATCTAGCGGTTGAAGG 642
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
   197 Asp-----AlaArgAspValTrpGlnIleSerMetThrHisIleAlaLeuGlnGly 213

```



## Alignment Scores:

Pred. No.:	5,21e-25	Length:	233
Score:	323.00	Matches:	85
Percent Similarity:	52.3%	Conservative:	39
Best Local Similarity:	35.9%	Mismatches:	81
Query Match:	17.7%	Indels:	32
DB:	7	Gaps:	8

US-09-751-299-3 (1-1014) x US-11-096-568A-6956 (1-233)

```
QY 265 GCCAAGCGGTTGGAAATCATGTCACCTTGGGATGATGATGACCGGCTGGGACACCTT 324
    |||  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::  ::
Db 2  AAlaGlySerTyrLysValHisLeuValMetGlyValIleGluAysArgGlyTyrThrLeu 21
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 325 TACATCAGTCAGTGGTTCATAGCGGATATGATGATGATGATGATGATGATGATGATG 384
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 22 TyrcysThrValLeuphepneaspserGlnGlyHisTyrLeuGlyLysHisArgLysIle 41
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 385 AAACCTACTTTTGTGAACGTACTTGTTCGGCGAAGGGATGGTTCATCGCGTTCGCGT 444
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 42 MetProthrAlaLeuGluArgValIletrpGlyPheGlyAspGlySerThrIleProVal 61
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 445 TTCGAGAGCTGCTTGGAAAGGCTGGGCTTATGCGTGGTGGAGACACTTCAACCGCTA 504
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 62 PheGluThrProValGlyLysIleGlyAlaIleCysStrpGluAsnArgMetProLeu 81
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 505 ACAAAATACGCTTGTATGACCAAAATGAAAGATTCATTTGCGGCTGGCCGAGCTTT 564
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 82 LeuArgThrAlaMetTyrAlaLysGlyValGluIleTyrCysAla----- 96
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 565 AGCCTTTATCCTAATGCGCGAAAGCCCTGGGGCTGATGATGATGATGATGATGATGATG 624
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 97 -----ProthrAlaasp-----AlaArgAspValIletrpGlnHisSerMetThr 110
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 625 ATCTAGCCGCTGGAAGGCAATGCTTCGTACTAGCGCTG-----TGTCG----- 669
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 111 HisIleAlaLeuGlnGlyGlyCysPheValLeuSerAlaAsnGlnPheCysArgArg 130
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 670 -----CTCGTTTCACAAATCCATGATGATGATGATGATGATGATGATGATGATG 705
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 131 AspTyrProProProGluTyrValPheAlaGlyThrGluValAspLeu-----Thr 148
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 706 GATGACGAAAGACATCGCTTGTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 765
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 149 ProAsp-----SerValValCysAlaGlyGly-----SerValIleIleSerPro 163
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 766 GATGCTGTGACTGTGCTCGCGCTCTTCCGCGAAATGAAAGGCTATTCCTACGCAAC 825
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 164 LeuGlyAlaValLeuAlaGlyProAsnTyrAspGly--GluAlaLeuIleSerAlaAsp 182
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 826 CTTCGATCCCTGGAGTACGATCCCTGCTAAATGCGCGAGACCTTCCTGCTGCTATTC 885
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 183 LeuAspLeuGlyGluIleAlaArgAlaLysPheAspPheAspValValGlyHisTyrSer 202
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 886 CGTCCGACATTAATCGCTTGTCTAATGATGACGCCCTTAATTACCGGTA 936
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 203 ArgProGluValLeuSerLeuThrValLysAspHisProThrAsnProVal 219
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
```

## RESULT 14

```
US-11-096-568A-20688
; Sequence 20688, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nickolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; TITLE OF INVENTION: Theby
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; PRIORITY FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 20688
; LENGTH: 193
; TYPE: PRT
```

ORGANISM: Zea mays subsp. mays

```
FEATURE:
NAME/KEY: misc_feature
LOCATION: (1)..(193)
OTHER INFORMATION: Ceres Seq. ID no. 12386687
US-11-096-568A-20688
```

## Alignment Scores:

Pred. No.:	4,92e-20	Length:	193
Score:	274.50	Matches:	68
Percent Similarity:	54.2%	Conservative:	36
Best Local Similarity:	35.4%	Mismatches:	65
Query Match:	15.0%	Indels:	23
DB:	7	Gaps:	7

US-09-751-299-3 (1-1014) x US-11-096-568A-20688 (1-193)

```
QY 388 CCTACTTTTGTGAACGTACTTGTTCGGCGAAGGGATGGTTCATCGCTACCGGTTTC 447
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 2  ProThrAlaLeuGluArgValPheTrpGlyPheGlyAspGlySerThrIleProValIle 21
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 448 GAGACGCTGCTTGGAAAGGCTGGGCTTATGCTGTTGGAGACACTTCAACCGCTTACA 507
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 22 AspThrProIleGlyLysMetGlyAlaLeuIleCysStrpGluAsnArgMetProLeu 41
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 508 AAATACGCTTGTATGACCAAAATGAAAGATTCATTTGCGGCTGGCCGAGCTTTAC 567
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 42 ArgThrAlaMetTyrAlaLysGlyIleGluIleTyrCysAla-----ProthrAlaAsp 59
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 568 CTTTATTCCTAATGCGCGAAAGCCCTGGGGCTGATGATGATGATGATGATGATGATG 627
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 60 CysMetProThrTrp-----LeuSerSerMetThrHis 70
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 628 TATGCCGTTGAAGGCAATGCTTCGTACTAGCGCTGCTGCTGCTTTCACATCATC 687
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 71 IleAlaLeuGlnGlyGlyCysPheValLeuSerAlaCysGlnPheCysArgArgAsn 90
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 688 ATCGAT-----ATGCTTGTACAGATGACGAAAG-----CATGCG 723
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 91 TyrProProProProGluTyrThrPheCysGlyLeuGlnGluIleGluProSerProGluSer 110
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 724 TTGCTTGTGCTGTGCTGTGACACTCAGTATCAAGAGCTGATGCTGCTGCTGCTGCTG 783
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 111 ValValCysSerGlyGly-----SerValIleIleSerProLeuGlyThrValLeuAla 128
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 784 GCGCCTCTTCCGCAAAATGAAAGAGGATATTCCTACGCAACCTGATCCCTGAGTACG 843
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 129 GlyProAsnTyrGluSer---GluAlaLeuThrAlaAspLeuAspLeuGlyIle 147
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 844 ATCCCTGCTAAATGCGCGAGACCTTCCTGCTGCTATTCCTCCGCGACATTATCGC 903
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 148 ValAlaGluAlaLysPheAspPheAspValValGlyHisTyrSerArgProGluValLeuSer 167
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
QY 904 TTGCTAATGATCGACGCCCTTAATTACCGGTACTT 939
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
Db 168 LeuValValLysSerAspProLys---ProAlaVal 178
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
```

## RESULT 15

```
US-11-079-463-5457
; Sequence 5457, Application US/11079463
; Publication No. US20060073161A1
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO BACTERIOIDES FRM
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: PATH00-03DIV2
; CURRENT APPLICATION NUMBER: US/11/079,463
; PRIORITY FILING DATE: 2005-03-14
; PRIOR APPLICATION NUMBER: US 60/128,705
; PRIOR FILING DATE: 1999-04-09
; PRIOR APPLICATION NUMBER: US 09/540,209
; PRIOR FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 10444
```

SEQ ID NO 5457  
 LENGTH: 295  
 TYPE: PRT  
 ORGANISM: B.fragilis  
 US-11-079-463-5457

Alignment Scores:  
 Pred. No.: 0.000105 Length: 295  
 Score: 126.00 Matches: 69  
 Percent Similarity: 38.9% Conservative: 40  
 Best Local Similarity: 24.6% Mismatches: 86  
 Query Match: 6.9% Indels: 85  
 DB: 7 Gaps: 14

US-09-751-299-3 (1-1014) x US-11-079-463-5457 (1-295)

QY 13 ATCAAGTGGCTGCGCAAGCCCGCATGATGATTGGAGCGAGGTGAC 72  
 Db 5 IleyValGlyIleIleGlnIaIa-----AsnThrSerAsp 17  
 QY 73 AAAACCATGAGTGTGATG-----GAGAGACGACGTAATATGCT 114  
 Db 18 IleArgIleAsnLeuMetAsnLeuAlaIylSerIleGluAlaCysAlaIaAsnGlyAla 37  
 QY 115 CGTCTGATGCGCTTCCGAAACTTGATTCAGGCTACCATGGTTCTTTGGCTTGAC 174  
 Db 38 HisLeuValIleLeuGlnIu-----LeuHis 46  
 QY 175 TCACCAAGCATGGCAATGCAATTTGTACGCCAATACATGAGAACTCATTTGAGATTG--- 231  
 Db 47 Asn-----SerLeuTyrPheCysGlnThrGluAsnThrAspLeuPheGluLeuAla 63  
 QY 232 -----GATGGCCCTCAAGCTAAGCCGATTTGAT-----GCAGCCAGCGGTTG 276  
 Db 64 GluProIleProGlyProSerThrGlyPheTyrSerGluLeuAlaIaIaAsnArgIle 83  
 QY 277 GGAATCATGATGCACCCCTGGGATGATGATGACCGGTCGGTGCACCCCTTACATCAGTCAG 336  
 Db 84 ---ValLeuValIleThrSerLeuPheGluIlyAsnGalaProGlyLeuTyrHisAsnThrAla 102  
 QY 337 TGGTTCATAGCGCATATGATGATGACACCATGGGCGCCGCAAGTTG----- 384  
 Db 103 ValValIlePheAspArgAspGlySerIleAlaGlyIylArgIylSerMetHisIleProAsp 122  
 QY 385 AAACCTACTTTGTTGAAACGTACTTTGTCGCGAAGGGAGTGTTCATCGCTACCGGTT 444  
 Db 123 AspProAlaTyrTyrGluIlyAspPheTyrPheThrProGlyAsp---IleGlyPheGluPro 141  
 QY 445 TTCGAGACGTCGTGGAGGCTGGGCTTATGCTGAGAGCACCTTCAACCGCTA 504  
 Db 142 IleGlnThrSerLeuGlyIylSerLeuValCysTrpAspGln----- 157  
 QY 505 ACAAAATPACGCTTTGTATGACAAAAATGAGATTCATTTGGCGCTTGCCGAGCTTT 564  
 Db 158 -----Tyr----- 158  
 QY 565 AGCCTTATCTTAATGCGCGAAACCCCTGGGCGCTGATGCAATGATAGCGGCTTCGA 624  
 Db 159 -----TyrProGluIaIaIa-----Arg 164  
 QY 625 ATCTATGCGGTTGAAGGCAATGCTTGCTACTAGCGTGTGCGCTGTTCAATCC 684  
 Db 165 LeuMetAlaLeuIylSerGlyAlaGluIleuIleTyrProThrAlaIleGlyTyrProIleSer 184  
 QY 685 ATGATCGATATGCTTTGATACAGATGACGAAAG-----CATGCGTTGCTTCTG 732  
 Db 185 -----ThrAspThrAspAspGluIylSerArgGlnLeuAsnAlaTripleIle 200  
 QY 733 GCTGATGATGACACTACGATATCATTAGGCGCTGATGCTGATGCTGCTGCGCGCTC 790  
 Db 201 SerGlnArgAlaHisAlaIaIaIaAsnGlyLeuProValIleSerValaAsnArgVal 219

Search completed: April 27, 2006, 01:52:58  
 Job time : 26.7372 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 26, 2006, 22:29:49 ; Search time 4775.42 Seconds  
(without alignments)  
9934.641 Million cell updates/sec

Title: US-09-751-299-3  
Perfect score: 1014  
Sequence: 1 atgaagaagatcatcaagtc.....cgcaactcgaagaatttga 1014

Scoring table: IDENTITY NUC  
Gapop 10.0 , Gapext 1.0

Searched: 41078325 seqs, 23393541228 residues

Total number of hits satisfying chosen parameters: 82156650

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :  
EST:  
1: gb\_est1:  
2: gb\_est2:  
3: gb\_est3:  
4: gb\_hic:  
5: gb\_est4:  
6: gb\_est5:  
7: gb\_est6:  
8: gb\_est7:  
9: gb\_gse1:  
10: gb\_gse2:  
11: gb\_gse3:

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	128	12.6	769	8	DR609823 EST999951
2	113.6	11.2	840	8	DR709497 Asn_11063
3	113	11.1	888	9	AQ329770 nbx50045P
4	105.6	10.4	551	7	CO149108 EST824161
5	105.6	10.4	553	7	CO149186 EST824239
6	105.6	10.4	610	7	CO134433
7	105.6	10.4	804	8	DR709485 Asn_11049
8	96.8	9.5	546	7	DR709485 Asn_11049
9	93.2	9.2	840	7	CN808140 BLOOD_EST
10	93.2	9.2	998	7	CN808249 BLOOD_EST
11	88.4	8.7	515	7	CO1337063 EST831734
12	87	8.6	655	6	CF453420 fts36_1
13	87	8.6	699	8	DR631467 EST102159
14	87	8.6	732	8	DR631867 EST102199
15	87	8.6	810	8	DR629486 EST101961
16	87	8.6	841	8	DR632130 EST102225
17	87	8.6	856	8	DR622046 EST101217
18	87	8.6	858	8	DR630522 EST102065
19	86.6	8.5	717	8	DN905823 31411.2 I
20	86.6	8.5	817	7	CK860468 31411.2 I
21	86.4	8.5	688	6	CA265056 SCARST305
22	86.4	8.5	694	6	CA283378 SCBGS105

23	86.4	8.5	840	7	CO170076	CO170076 ND1_11_C
24	86.4	8.5	855	8	DR163242	DR163242 RTFBI_41
25	86.2	8.5	622	9	AQ329769	AQ329769 nbx50045P
26	85.4	8.4	695	6	CA185867	CA185867 SCSPST307
27	84.2	8.3	705	8	DR523243	DR523243 MS02714_B
28	83.4	8.2	810	2	BG128900	BG128900 EST474546
29	83.4	8.2	728	3	BG123280	BG123280 EST468926
30	83.4	8.2	736	3	BI925888	BI925888 EST545777
31	82.6	8.1	815	7	CN136550	CN136550 OX1_43_HO
32	82.2	8.1	701	2	BE660909	BE660909 1155_Gmax
33	82.2	8.1	895	7	CO163629	CO163629 FLDI_42_G
34	81.8	8.1	607	7	CO364093	CO364093 RTK1_13_D
35	81.8	8.1	660	6	CD717845	CD717845 VVB143F06
36	81.8	8.1	668	6	CD715152	CD715152 VVB208G06
37	81.8	8.1	642	6	CB342297	CB342297 CA32EN000
38	81.6	8.0	690	3	BI418062	BI418062 LjNEST24b
39	81.6	8.0	1087	1	AM720658	AM720658 LjNEST16h
40	81	8.0	789	7	CN133019	CN133019 OX1_9_D07
41	81	8.0	792	7	CN131548	CN131548 RHOH1_48
42	81	8.0	797	7	CN124609	CN124609 RHOH1_5_G
43	81	8.0	802	7	CN126484	CN126484 RHOH1_17
44	81	8.0	833	7	CN130894	CN130894 RHOH1_44
45	80.2	7.9	902	8	DR525110	DR525110 WS02719_B

#### ALIGNMENTS

RESULT 1  
LOCUS DR609823 769 bp mRNA linear EST 11-JUL-2005  
DEFINITION EST999951 FVG Gibberella moniliformis cDNA clone FVGB83, mRNA  
ACCESSION DR609823 GI:70684471  
VERSION DR609823.1  
KEYWORDS  
SOURCE Gibberella moniliformis  
ORGANISM Gibberella moniliformis  
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes; Hypocreomycetidae; Hypocreales; Nectriaceae; Gibberella.  
REFERENCE Brown, D.W., Cheung, F., Proctor, R.H., Butcho, A.E., Zheng, L., Lee, Y., Kendera, D.F., Smith, S., Feldblum, T., Glenn, A.E., Plattner, R.D.,  
Kendra, D.F., Town, C.D., and Whitelyaw, C.A.  
Analysis of 87,000 expressed sequence tags reveals alternatively spliced introns in multiple genes of the fumonisin gene cluster  
Unpublished (2005)  
JOURNAL  
COMMENT USDA/ARS/NCAR  
Contact: Brown, D.W.  
USDA  
1815 N. University St, Peoria, IL 61604, USA  
Tel: 309 681 6230  
Fax: 309 681 6689  
Email: brown@ncar.usda.gov  
TIGR sequence name: FVGB83TH  
Seq primer: AAT TAA CCC TCA AAG GG.  
Location/Qualifiers  
1. 769  
/organism="Gibberella moniliformis"  
/mol\_type="mRNA"  
/strain="m125"  
/db\_xref="taxon:117187"  
/clone\_type="FVGB83"  
/clone\_type="mycelia"  
/clone\_lib="FVG"  
/note="Vector: pBluescript II SK(+); Site 1: EcoRI;  
Site 2: XhoI; anamorph: Fusarium verticillioides. Mycelia  
was collected after growth in liquid GYM medium for 96  
hours. Cultures were vacuum filtered and the mycelial mats  
were frozen in liquid nitrogen, ground to a powder, and  
then added to Trizol Reagent (Invitrogen, Carlsbad CA) at  
approximately 1 g mycelia per 10 ml Trizol. The cDNA was  
directionally ligated into the pBluescript II SK(+) XR

#### FEATURES

source

ORIGIN vector (cDNA Synthesis Kit; Stratagene). "

Query Match 12.6%; Score 128; DB 8; Length 769;  
Best Local Similarity 53.0%; Pred. No. 1,1e-28;  
Matches 297; Conservative 0; Mismatches 260; Indels 3; Gaps 1;

QY 1 ATGAAAGAGCTATCAGAGTGGCTGCTGCAAGCCCGCCGATCTACATGATTTGGAG 60  
DB 43 ATGTCCAGTCCCTCAAGGTGCGCCCATTCACAGCCGCCGCTTGAAAGATCTCCAG 102  
QY 61 GCGACGGGCAACAACCATGAGTGTATGAGAAAGACGCAATTAATGCTGCTG 120  
DB 103 GTGGCGTCAACAAGTCCATGCTGTCTATCCAAAGGACGAAAGAGGGTCCAGCTC 162  
QY 121 ATTCCTTTCCGGAACCTTGATTCAGGCTAACCCATGGTTCCTTTGGCTTACCTCA 180  
DB 163 ATTTGGCTTCCCTGAAGTCTTATCCCTGGATATTCATGAGCATCTGGGCCAAGCTG 222  
QY 181 GCATGGCGAATGCAATTTGTACGCCAATACCATGAACTCATTTGAGTTGAGCCCT 240  
DB 223 ACCGAGAACGACCATGATCAATGATGATTTCAAGAACTCAATGAGAAAGATGACCT 282  
QY 241 CAAGCTAAGCCGATTTGATGAGCAGCCAGCGGTTGGGAATCATGTCACCTGGGGATG 300  
DB 283 GAGATGACCAAGATCCAGCTGCTGTTCGAGAGGAGGTGCTTTGTATGCTCTTGATAC 342  
QY 301 AGTGAACGGGTGCTGGGACCCCTTTACATCAATGATGCTTCAATGAGGATATGCTGAC 360  
DB 343 AGCGAGATATCAGGGGAAACCTTTACATCCGACAGTCTTTCATGACGAGACCGGCACT 402  
QY 361 ACCATTTGGGGCCCGGCGAAAGTTGAACCTATTGTTGAACGTAATCTTTGTTGGCGAA 420  
DB 403 ATTTGTTCCACCGTGCAGAAATCAAGCTTACCCACGTTGAGCGGTATCTACCGTAC 462  
QY 421 GGGGATGTTATGCTAGCTAGC---GGTTTTCGAGAGCTGTGTTGAAGCTGGGTGCTTGA 477  
DB 463 GGACAGGGCGAGTCTCTGACCAAGCTCGCCGACGAAAGTTGGAAGGGTGTGCTGCTT 522  
QY 478 TGTGTTGGAGACCTTCAACCGCTTAACAAATACGCTTGTATGACAAATGAGAG 537  
DB 523 AACTGCTGGGAGCAACCCAGACACTTCTCCGCTACTACGATACCTCCAGAGACCTGAT 582  
QY 538 ATTGATTTGGCGGCTTGGCC 557  
DB 583 ATCCAGTCTCCAGCTGGCC 602

RESULT 2  
LOCUS DR709497 840 bp mRNA linear EST 14-JUL-2005  
DEFINITION Asn\_11063 *Aspergillus niger* pBluescript (EcoRI-XhoI) *Aspergillus niger* cDNA clone Asn\_11063, mRNA sequence.  
ACCESSION DR709497  
VERSION DR709497.1 GI:70825788  
KEYWORDS EST.  
SOURCE *Aspergillus niger*  
ORGANISM *Aspergillus niger*  
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes; Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; *Aspergillus*.  
REFERENCE 1 (bases 1 to 840)  
AUTHORS Tsang A., Storms R. and Butler G.  
TITLE Expressed sequence tags from *Aspergillus niger* cDNA library  
JOURNAL Unpublished (2005)  
COMMENT Contact: Tsang A  
Centre for Structural and Functional Genomics  
Concordia University  
7141 Sherbrooke Street West, SP545-1, Montreal, QC H4B1R6, Canada  
Tel: 514 848 2424 3405  
Fax: 514 848 4504  
Email: tsang@vax2.concordia.ca  
POLYA=No.  
Location/Qualifiers

ORIGIN source

1..840  
/organism="Aspergillus niger"  
/mol\_type="mRNA"  
/strain="N402"  
/db\_xref="taxon:5061"  
/clone="Asn\_11063"  
/dev\_stage="mycelial growth"  
/lab\_host="E. coli"  
/clone\_lib="Aspergillus niger pBluescript (EcoRI-XhoI)"  
/note="Vector: pBluescript KS+; Site\_1: XhoI; Site\_2: EcoRI; Complementary DNA was synthesized with ZAP Kit (Stratagene) using poly(A)+RNA isolated from *Aspergillus niger* cultured under different carbon sources (glucose, maltose, xylose, lactose, sorbitol, xylan, and bran). Synthesis was primed with oligo(dT)/XhoI primer. EcoRI adaptors were ligated to the blunt-ended, double-stranded cDNA. The EcoRI-XhoI-digested cDNA was ligated with EcoRI-XhoI-digested pBluescript KS+ (Invitrogen Corp). "

Query Match 11.2%; Score 113.6; DB 8; Length 840;  
Best Local Similarity 49.6%; Pred. No. 4,1e-24;  
Matches 411; Conservative 0; Mismatches 399; Indels 18; Gaps 4;

QY 19 GTGCGCTGCGGCAAGCCGCCGATCTACATGATTTGGAGCGGACCAAAACC 78  
DB 22 GTCCGACGACCCAGGACGAGCCAGTCTGCGTGAATTGGAAGCAACCGTAAAGACA 81  
QY 79 ATTAGATTGAGAAAGACGACGATATATGCTCTGATGCGCTTTCCGAAACT 138  
DB 82 TGTGACCTTATTTGCTGAAGCAGCTGCGAACCGCCTCAGTTGTGTGATTTCCGAGTGT 141  
QY 139 TGGATTCAGGCTTACCATGATGTTTCTTTGGCTTATCACCAGATGGGCAATGATTT 198  
DB 142 TGGATCCCGGAAATCTGCTGGATTTG-----GACGCGCTTTGACATGCGCTTA 195  
QY 199 GTACGCCAATATCATGAGAACTCATTTGAGTTGAGGCCCTCAAGCTTAAGCGATTTCA 258  
DB 196 TCACTATTTATATCAAAAATCTCCCTGAAGTTATTTCCCGCAATAGGCCAGATCCAG 255  
QY 259 GATGACGCAAGCGGTTGGGAATCATGTTCACCTCGGGAGATGAGTGAACGGGTGGTGGC 318  
DB 256 CAATGCGGACGGGAAATTAATCGTGGTGTGTGTGGCTTTTCCGAAACCTGCATTAAC 315  
QY 319 ACCCTTACATCATGATGCTGTTGATAGGCGATATGAGTGAACCATTTGGGGCCGGGCA 378  
DB 316 TCCCTGATATCTCGAGGCTATTAATTTGCAAGTATGAGGAAGATCTCACACCCGCAAG 375  
QY 379 AAGTTGAACCTACTTTTGTGAACGTACTTTGTTCCGCGAAGGGAGTTCATCGCTA 438  
DB 376 AAAATCAAGCCACATCATATGAGAGGACCAATTTTCGGTGACTGTTTGGAGACTGCTG 435  
QY 439 GCG---GTTTTCGAGAGCTCTGTTGGAAGGCTGGGTGCTTATGCTGTTGGAGACCTT 495  
DB 436 CAGAGTGAATGCTATCATCAAGCCGCTGCGCTTGTCTCTTCTGCTGGAGATATC 495  
QY 496 CAACGCTTAACAAATATAGCTTTGATGACAAATATGAAGATTCATTTGGCGCTTGG 555  
DB 496 CAGCTCTGCTCAAGTATCAACTTATGCGAGCGGGAACAAATTCATGTTGCCGCAATG 555  
QY 556 CCGAGCTTTAAGCTTTATCTTAA-----TSCGGGAAAGCCCTGGGGCTGATGTCAT 609  
DB 556 CCGCTCTGTTCCCATAGGAGGAGCGTTCTTGTGTTTGCATGATGTCACGAGGGGACCC 615  
QY 610 GTAGGCGCTCTCGAATTTATGCGGTTAAAGGCAATGCTTCGTAAGCGTCTGTGCG 669  
DB 616 AGTTCCATTTGAGAGGACCTATGCAATGCAAGTCAAGTCTTTGTGCTTCATACCAACA 675  
QY 670 CTGCTTCAACATCATGATGATATGCTTTGTATGACATGACGAAAGCATGCTTGTGCT 729  
DB 676 GTCATAGGCAAGTCTGGCATTTGACCGGATGGCCACATGACTGGGGCCC---TATATAGC 732  
QY 730 CTGCGTGGTGTGACACTCAAGTATCATAGGGCCCTGATGTTGTGACTTGGTGGCGCCT 789



Db 733 ACCCTGGCGAGGTGCTCCGCCATCTTCGGGCCCGCAGCGCTGCATTTTCGCAACC 792  
Oy 790 CTTGCCGAATGAAGAGGTATTCTTACCGCAACCTTGATCTTGA 837  
Db 793 ATTCTATGCAAGAGGAGGATTATCAACCCGACCTCGACTTTGAA 840

RESULT 3  
AQ329770/c 888 bp DNA linear GSS 08-JAN-1999

LOCUS nbdb0045P09r CUGI Rice BAC Library Oryza sativa (japonica  
DEFINITION cultivar-group) genomic clone nbdb0045P09r, genomic survey  
sequence.

ACCESSION AQ329770.1 GI:4121620  
VERSION AQ329770.1  
KEYWORDS GSS.  
SOURCE Oryza sativa (japonica cultivar-group)  
ORGANISM Oryza sativa (japonica cultivar-group)  
Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;  
Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae;  
Ehartoideae; Oryzaceae; Oryza.

REFERENCE 1 (bases 1 to 888)  
Wing, R.A. and Dean, R.A.

A BAC End Sequencing Framework to Sequence the Rice Genome  
Unpublished (1998)  
Contact: Wing RA  
Clemson University Genomics Institute  
Clemson University  
100 Jordan Hall, Clemson, SC 29634, USA  
Tel: 864 656 7288  
Fax: 864 656 4293  
Email: rwing@clemson.edu  
Seg primer: GGAACAGCTATGACCATG  
Class: BAC ends  
High quality sequence start: 13  
High quality sequence stop: 225.  
Location/Qualifiers

FEATURES  
source 1..888  
/organism="Oryza sativa (japonica cultivar-group)"  
/mol\_type="genomic DNA"  
/cultivar="japonica"  
/cultivar="Nipponbare"  
/db\_xref="taxon:39947"  
/clone="nbdb0045P09r"  
/tissue\_type="leaf"  
/lab\_host="E. coli DH10B"  
/clone\_lib="CUGI Rice BAC library"  
/note="Vector: pBelorAC11, Site 1: HindIII, Site 2:  
HindIII; Rice is one of two most popular grains in the  
world. Half of the world population especially those  
inhabiting highly populated areas of the humid tropics  
and subtropics, rely on rice as their primary source of  
carbohydrate. Monocotyledonous rice is a diploid plant  
(2n=24) with a haploid genome equivalent of 431 Mbp  
(Arumuganathan and Earle, 1991). The relatively small  
genome of rice, three times larger than that of  
Arabidopsis, makes it suitable for genomic studies. In  
order to facilitate positional cloning, physical mapping  
and genome sequencing of rice, we have constructed a BAC  
library from Oryza sativa, Nipponbare variety. The  
library contains 36,864 clones with an average insert size  
of 128.5 Kb providing 10.9 haploid genome equivalents. The  
deep coverage allows the isolation a particular sequence  
with a probability of 99.9 %. Two high density filters,  
each containing 18,432 clones (doubly spotted), represent  
the whole library for colony screening."

## ORIGIN

Query Match 11.1%; Score 113; DB 9; Length 888;  
Best Local Similarity 56.2%; Pred. No. 6,5e-24;  
Matches 209; Conservative 0; Mismatches 163; Indels 0; Gaps 0;

Oy 183 ATGGCAATGCAATTTTGTACCCCAATACCATGAGAATCATGTGATGGCCCTCA 242  
Db 372 ATGGGGAAGAGATTTAGATCATGACAGTCAAGAAAACAGANNGGCCGATGGCCGGA 313  
Oy 243 AGCTAAGGCAATTTTCAATGACAGCCAAAGCGGTGGAAATCATGTCACCTGGGGATAG 302  
Db 312 GTCCGCGGNATTCGCCGAGGCCGCGCTCGGCACGCAATGTGATCATGTCTGGGCTATAG 253  
Oy 303 TGAACGGGTGGTGGACACCTTTTATCATGACATGTGTTCTAGCCGATATGTGGACAC 362  
Db 252 CGAAGCGAGCCGCGGAGCCCTCTACATGCTCACTGATGATGATGCGGACCAT 193  
Oy 363 CATTTGGGCGCCGGGAAAGTTGAACCTTATTTTGAAGATCTTTTGGGGAAG 422  
Db 192 CGTACAGCGCGGAGCAAGCATCAAGCAACCAATGTGAGGCGACATCTTCGTGATGG 133  
Oy 423 GGATGTTTCATGCTTACGCGTTTTCAGACCTGTGTGAAAGCTGGGTGCTTATGCTG 482  
Db 132 CGACGCGAGCGACATCCGCGTGTGTGAGGACCTTGGCCGCTTGGATGCTTGTCTG 73  
Oy 483 TTGGAGACCTTTCACCGCTACCAAAATATGCTTTGTATGACAAATGAAGATTCA 542  
Db 72 CTGGGAACATATGAACCCGCTCACCAATATGCAATGATGACGCCAGACGACAGATCA 13  
Oy 543 TTGTGCGGCTTG 554  
Db 12 TCTGCGGCATG 1

RESULT 4  
COL49108 551 bp mRNA linear EST 17-JUN-2004  
LOCUS EST824161 Aspergillus flavus Normalized cDNA Expression Library  
DEFINITION Aspergillus flavus CDNA NAGDP49 5' end, mRNA sequence.  
ACCESSION COL49108  
VERSION COL49108.1 GI:48903109  
KEYWORDS EST.  
SOURCE Aspergillus flavus  
ORGANISM Aspergillus flavus  
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.  
1 (bases 1 to 551)  
REFERENCE Yu, J., Whitelaw, C.A., Nierman, W.C., Bhatnagar, D. and Cleveland, T.E.  
Aspergillus flavus expressed sequence tags for identification of  
genes with putative roles in aflatoxin contamination of crops  
FEMS Microbiol. Lett. (2004) In press  
JOURNAL Contact: Yu J  
COMMENT Food and Feed Safety Research Unit  
USDA/ARS, Southern Regional Research Center  
1100 Robert E. Lee Boulevard, New Orleans, LA 70124, USA  
Tel: 504 286 4405  
Fax: 504 286 4419  
Email: jjuyu@ars.ars.usda.gov  
Contact Dr. Yu at USDA/ARS SRRC (jjuyu@ars.ars.usda.gov) for clone  
information  
PCR Primers  
FORWARD: M13F  
BACKWARD: M13R  
Seq primer: M13 Forward  
POLYA=No.

FEATURES  
source 1..551  
Location/Qualifiers

/organism="Aspergillus flavus"  
/mol\_type="mRNA"  
/strain="NRRL 3357"  
/db\_xref="taxon:5059"  
/clone="NAGDP49"  
/sex="asexual mycelia"  
/cell\_type="mycelia"  
/dev\_stage="developmental stages from 18 to 96 hours"  
/lab\_host="E. coli DH10B T1 resistant cells"  
/clone\_lib="Aspergillus flavus Normalized cDNA Expression  
Library"



RESULT 6  
CO134433  
LOCUS  
DEFINITION EST229104 Aspergillus flavus Normalized cDNA Expression Library  
ACCESSION CO134433  
VERSION CO134433.1 GI:48883411  
KEYWORDS EST.  
SOURCE Aspergillus flavus  
ORGANISM Aspergillus flavus  
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
Fusiculales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.  
1 (bases 1 to 610)  
Yu.J., Whitelaw,C.A., Nieman,W.C., Bhatnagar,D. and Cleveland,T.E.  
Aspergillus flavus expressed sequence tags for identification of  
genes with putative roles in aflatoxin contamination of crops  
FEMS Microbiol. Lett. (2004) In press  
JOURNAL  
COMMENT Contact: Yu J  
Food and Feed Safety Research Unit  
USDA/ARS, Southern Regional Research Center  
1100 Robert E. Lee Boulevard, New Orleans, LA 70124, USA  
Tel: 504 286 4405  
Fax: 504 286 4419  
Email: jinyu@arcc.ars.usda.gov  
Contact Dr. Yu at USDA/ARS SRRC (jinyu@arcc.ars.usda.gov) for clone  
information  
PCR Primers  
FORWARD: M13F  
BACKWARD: M13R  
Seq primer: M13 Forward  
POLYA=No.

FEATURES  
source  
location/Qualifiers  
1..610  
/organism="Aspergillus flavus"  
/mol\_type="mRNA"  
/strain="NRRL 3357"  
/db\_xref="taxon:5059"  
/clone="NAFAS28"  
/sex="asexual mycelia"  
/cell\_type="mycelia"  
/dev\_stage="developmental stages from 18 to 96 hours"  
/lab\_host="E. coli DH10B TI resistant cells"  
/clone\_lib="Aspergillus flavus Normalized cDNA Expression  
Library"  
/note="Vector: pBluescript (SK+) (Stratagene), antibiotic  
selection marker: Carbenicillin, Site\_1: NotI, at the 5  
prime end; Site\_2: EcoRI, at the 3 prime end; This  
normalized cDNA expression library was constructed using a  
mixture of mycelial cells grown under eight different  
medium conditions and harvested at 5 time points (18, 24,  
48, 72, 96 hours). The poly-A sequence was trimmed off  
before ligating to vector."

ORIGIN  
Query Match 10.4%; Score 105.6; DB 7; Length 610;  
Best Local Similarity 51.5%; Pred. No. 1,3e-21;  
Matches 268; Conservative 0; Mismatches 249; Indels 3; Gaps 1;

QY 14 TCAGAGTGGCTGCTGTCGACGCGCCCGCATCTACATGATGATTTGGAGCGACGATGACA 73  
DB 91 TCAAGTTGCTGCGGTTCAAGCTGAGCTGTCGTGGAATGACCTACAAGCGCGGTTTCCA 150  
QY 74 AAACCATGAGTTGATGGAAGACGACGATTAATGCTGCTGTGATCGCTTTCCGG 133  
DB 151 AGGTATCTCACTGATCAAGGACGCTGCCAATAAAGCGGCCAAGGTGTCGTTTCCCG 210  
QY 134 AAACCTGGATTCAGGCTACCATGATGTTTCTTTGGCTTGACACACAGATGGGAATGC 193  
DB 211 AGGTATTCATCCCTGTTATTCATGAGATCTGAGACTTAATCGGCTATTTGAAATGTTT 270  
QY 194 AATTGTACGCGCAATACATGAGAACTGATGAGATGGATGGCCCTCAAGCTAAGCCCA 253  
DB 271 CTTATGAAAGAGATCTTCGAACTCATGGAAGGAATCGAAGAGATGAGAACCA 330

QY 254 TTTCAGATGACGACCAAGCGCTTGGGATCATNGTCAACCTGGGGATGATGAACGGGTGC 313  
DB 331 TCAAAATATGCTCCGTGGCGAAGCGGTGATTCATGTTCTTAGGATGATGAACGGTACC 390  
QY 314 GTGCAACCCCTTACATCAGTCAAGTGTTCATGAGCGATATATGAGACACCATTTGGGGCC 373  
DB 391 GAGGATCATTTGACATTTGCTCATTTATTCATTCCTACTACTGATCTATCTGCACACAC 450  
QY 374 GCGGAAGTTGAACCTTCTTTGTTGAACGATCTTTGTCGGCGAAGGGATGTTTCAT 433  
DB 451 GACGGAATATCAAGCGCACTCATGTCGAGAAGCTATTGGGAGACGTCAGACGACT 510  
QY 434 CGCT--AGCGTTTTCAGACGTTGTTGGAAAGCTGGGTGCTTATGCTTTGGAGC 490  
DB 511 CTTTAAAGACGTTAGCCCGCCAGTGTCTTTTGGAAACATTTGGGGGCTCACTGTTGGAGC 570  
QY 491 ACCTTCAACCGCTAACAAATACGCTTTGTATGACAAATA 530  
DB 571 ATACCAACCACTTCTACATGATTTACGATATGCTCAAAA 610

RESULT 7  
DR709485 804 bp mRNA linear EST 14-JUL-2005  
LOCUS  
DEFINITION Asn\_11049 Aspergillus niger pBluescript (EcoRI-XhoI) Aspergillus  
niger cDNA clone Asn\_11049, mRNA sequence.  
ACCESSION DR709485  
VERSION DR709485.1 GI:70825776  
KEYWORDS EST.  
SOURCE Aspergillus niger  
ORGANISM Aspergillus niger  
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
Fusiculales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.  
1 (bases 1 to 804)  
Tsang,A., Storms,R. and Bulter,G.  
Expressed sequence tags from Aspergillus niger cDNA library  
Unpublished (2005)  
JOURNAL  
COMMENT Contact: Tsang A  
Centre for Structural and Functional Genomics  
Concordia University  
7141 Sherbrooke Street West, SP545-1, Montreal, QC H4B1R6, Canada  
Tel: 514 848 2424 3405  
Fax: 514 848 4504  
Email: tsang@vax2.concordia.ca  
POLYA=No.

FEATURES  
source  
location/Qualifiers  
1..804  
/organism="Aspergillus niger"  
/mol\_type="mRNA"  
/strain="NA402"  
/db\_xref="taxon:5061"  
/clone="Asn\_11049"  
/dev\_stage="mycelial growth"  
/lab\_host="E. coli"  
/clone\_lib="Aspergillus niger pBluescript (EcoRI-XhoI)"  
/note="Vector: pBluescript KS+; Site\_1: XhoI; Site\_2:  
EcoRI; Complementary DNA was synthesized with ZAP kit  
(Stratagene) using poly(A)+RNA isolated from Aspergillus  
niger cultured under different carbon sources (glucose,  
maltose, xylose, lactose, sorbitol, xylan, and bran).  
Synthesis was primed with oligo(dT)/XhoI primer. EcoRI  
adaptors were ligated to the blunt-ended, double-stranded  
cDNA. The EcoRI-XhoI-digested cDNA was ligated with  
EcoRI-XhoI-digested pBluescript KS+ (Invitrogen Corp)."

ORIGIN  
Query Match 10.4%; Score 105.6; DB 8; Length 804;  
Best Local Similarity 52.0%; Pred. No. 1.4e-21;  
Matches 290; Conservative 0; Mismatches 259; Indels 9; Gaps 2;

QY 19 GTCCCTGCGGCAAGCGCGCCGATCTACATGATTTGGAGGAGCGGTGACAAAAC 78

Db 23 GTGCGACTCACCCAGGACAGCCAGTGTGGCTGGACTTGGAGAACCCGTGAAAAAGACA 82  
Qy 79 ATTGAGTTGATGGAAGAAGACAGCATTAATAGCTCTGTGATCGCTTTCCGGAAAC 138  
Db 83 TGTACACTTATTTGCTGAAGAGCTGCGAACCGGCTCAGTTGGACTTTTCCGAGTGT 142  
Qy 139 TGAATTCAGGCTACCCATGGTTCTTTGGCTGACTCACGACATGGGCAATGCAATTT 198  
Db 143 TGGATCCCGGATATTC-----TGCCTGATTTGGGACAGGCTGTGACATGCGCCTA 196  
Qy 199 GTACGCCAATACCATGAGAACTTGGAGTTGAGTGAGCCCTCAAGCTAAGCGATTTCA 258  
Db 197 TCATCTATTATTCATCAAAAACCTCCCTGAAGATGATTCCTCCGCAAAATGCGCAGATCAG 256  
Qy 259 GATGACGCCAAGCGGTTGGGAATCATGTGTACCCGCGGAGATGAGTGAACGGGTGCGTGC 318  
Db 257 CAATGCGAGCGGAATAAATCGTGTGTGTGTGCTTTCCGAGAACTTCGATTAAC 316  
Qy 319 ACCCTTACATCATGTCAGTGTTCATAGCGATTAATGTTGACACCAATTGGGGCCCGGCGA 378  
Db 317 TCCCTGATATTCGACAGGCTATTATTCAGATTATGGAAAGATCCTCACCAACCGCGAAG 376  
Qy 379 AAGTTGAACCTACTTTTGTGAAGCTATTGTTGGCGAAGGAGGATGTTCAATCGCTA 438  
Db 377 AAAATCAAGCCCATCATATGAGCGGACATTTTCGGTGACTCGTTTGGAGACTGTCTG 436  
Qy 439 GGG---GTTTTCAGAGCTGTGTGAAAGGCTGAGGCTTAATGCTGTTGGAGACCTT 495  
Db 437 CAGAGTGATGATGATATCATACAGCCGGTGGCTGCTTCCGTGCGGAGCATATC 496  
Qy 496 CAACCGCTAACAATAATCGCTTTGATGACAAAATGAAGATCATTTGTCGGCTTGG 555  
Db 497 CAGCTCTCTCTCAAGATGACACTTAATGCGAGCGGGAACAATTCATGTTCCGATAG 556  
Qy 556 CCGAGCTTAGCCTTAT 573  
Db 557 CCGCTCTGTCCCTCAT 574

RESULT 8  
LOCUS CO139118 546 bp mRNA linear EST 17-JUN-2004  
DEFINITION EST833789 Aspergillus flavus Normalized cDNA Expression Library  
ACCESSION CO139118  
VERSION CO139118  
KEYWORDS GI:48888096  
SOURCE EST.  
ORGANISM Aspergillus flavus  
Aspergillus flavus  
Eukaryotes; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.  
1 (bases 1 to 546)  
Yu, J., Whitelaw, C.A., Nieman, W.C., Bhatnagar, D. and Cleveland, T.E.  
Aspergillus flavus expressed sequence tags for identification of  
genes with putative roles in aflatoxin contamination of crops  
FEMS Microbiol. Lett. (2004) In press  
CONTACT: Yu J  
Food and Feed Safety Research Unit  
USDA/ARS, Southern Regional Research Center  
1100 Robert E. Lee Boulevard, New Orleans, LA 70124, USA  
Tel: 504 286 4405  
Fax: 504 286 4419  
Email: jiyu@ars.ars.usda.gov  
Contact Dr. Yu at USDA/ARS SRRC (jiyu@ars.ars.usda.gov) for clone  
information  
PCR primers  
FORWARD: M13F  
BACKWARD: M13R  
Seq primer: M13 Forward  
POLYA=No.  
Location/Qualifiers  
1..546  
/organism="Aspergillus flavus"

## ORIGIN

Query Match 9.5%; Score 96.8; DB 7; Length 546;  
Best Local Similarity 51.2%; Pred. No. 7, 8e-19;  
Matches 252; Conservative 0; Mismatches 237; Indels 3; Gaps 1;

/mol\_type="mRNA"  
/strain="NRRL 3357"  
/db\_xref="taxon:5059"  
/clone="NAFDB72"  
/sex="asexual mycelia"  
/cell\_type="mycelia"  
/dev\_stage="developmental stages from 18 to 96 hours"  
/lab\_host="B. coli DH10B T1 resistant cells"  
/clone\_lib="Aspergillus flavus Normalized cDNA Expression  
Library"  
/note="Vector: pBluescript (SK+) (Stratagene), antibiotic  
selection marker: Carbenicillin; Site\_1: NotI, at the 5  
prime end; Site\_2: EcoRI, at the 3 prime end; This  
normalized cDNA expression library was constructed using a  
mixture of mycelial cells grown under eight different  
medium conditions and harvested at 5 time points (18, 24,  
48, 72, 96 hours). The poly-A sequence was trimmed off  
before ligating to vector."

RESULT 9  
LOCUS CN808140 840 bp mRNA linear EST 27-MAY-2004  
DEFINITION Blood EST0223 Metarhizium anisopliae ARSEF 2575 from insect blood  
Metarhizium anisopliae cDNA clone B602 5', mRNA sequence.  
ACCESSION CN808140  
VERSION CN808140.1 GI:47729613  
KEYWORDS EST.  
SOURCE Metarhizium anisopliae  
ORGANISM Metarhizium anisopliae  
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
Hypocreomycetidae; Hypocreales; Clavicipitaceae; mitosporic

REFERENCE 1 (bases 1 to 840)  
AUTHORS Wang, C.S., Hu, G. and St. Leger, R.J.  
TITLE Gene expression profiling of *Metarhizium anisopliae* grown under different conditions: mechanisms of fungal opportunism  
JOURNAL Unpublished (2004)  
COMMENT Contact: Wang CS  
Department of Entomology  
University of Maryland  
4112 Plant Sciences Building, College Park, MD 20742, USA  
Email: cwang4@umd.edu  
Seq primer: M13 Reverse.

FEATURES  
source 1..840  
Location/Qualifiers  
/organism="Metarhizium anisopliae"  
/mol\_type="mRNA"  
/db\_xref="taxon:5530"  
/clone="B602"  
blood\_lib="Metarhizium anisopliae ARSEF 2575 from insect blood"  
/note="Vector: PCMV.SPORT6.1; Metarhizium anisopliae was grown in insect haemolymph for 24 hours. A cDNA library was constructed in the vector PCMV.SPORT6.1"

## ORIGIN

Query Match 9.2%; Score 93.2; DB 7; Length 840;  
Best Local Similarity 49.5%; Pred. No. 1.3e-17;  
Matches 270; Conservative 0; Mismatches 273; Indels 3; Gaps 1;

```
QY 15 CAAGGTCGCTGCGTCAAGCCGCCGATCTACATGATTTGGAGCGACGGTGGACAA 74
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 84 CAAGGTCGCTGCGTCAAGCCGCCGATCTACATGATTTGGAGCGACGGTGGACAA 143
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 75 AACATGAGTTGATGAGAAAGACGACAGTAATTAATGCTCGTCTGATGCCCTTCCGA 134
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 144 GACTGTGACTGATCTCGAAGCTGGCGAAGATGTCAAGCTTATGCTTCTCTGA 203
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 135 AACTTGATTCAGGTCACCATGCTTTCTTGGCTTGACTCAGCAGCATGGGCATGCA 194
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 204 AGTTTGATTCCTGATTTCCCTATTGGCTGGCGTGTCACTACCAAGATTTCGCTGCC 263
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 195 ATTTGAGCCCAATACATGAGAACTCATTTGAGTTGATGGCCCTCAAGCTAAGCGCAT 254
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 264 CCTCTCAAGAAAGTTTACACAGAACGATCGGCCGACTGGAGAAATGCGACGAAT 323
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 255 TTCAAGTCAGCCAAAGCGGTGGGAATCATGTGTCACCTTGGGAGTGAAGAAAGGTCGG 314
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 324 TCGCAGGCTGCGCAAGCGGCTAAGATCTACGCTCTGTTGGGATATTCGAGCTGACGG 383
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 315 TGGCACCTTTACATCATGCTAGTGGTTCAATAGCGCATTAATGTGACACCATTTGGGGCCG 374
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 384 CCATACCATGTATCATGCGCAGATCATTCATGATCCACCGGAACCGTATCAACCAACCG 443
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 375 GGGAAAGTTGAAACCTACTTTTGTGAAAGTACTTTGTCGGGGAAGGGGATGGTTCAATC 434
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 444 TGCAGATCAAGCCCAACCATGTCGAAAGCTTGTATTTGGGAGGGAAGCGGTGACTC 503
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 435 GCT---AGCGTTTTCAGACGTCGTGTGAAAGGCTGGGTGGCTTATGCTGTTGGAGCA 491
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 504 TCTTCAAGCGCGTATGAGACTGAATTTGGCAATCTCGGCCACCTGAACTGGGGAATA 563
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 492 CCTTCAACCGCTAACAAATACGCTTTGTATGCACAAAATGAAGATTCATTGGCGGC 551
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 564 TATGAACCCCTTCTCAAGCGCTCAACCGCAGCAAAACGAAAGAAATCAGATAGCCGC 623
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 552 TTGGCC 557
   |||||
DB 624 ATGGCC 629
   |||||
```

RESULT 10  
CN808249 998 bp mRNA linear EST 27-MAY-2004  
LOCUS CN808249

DEFINITION Blood EST0332 *Metarhizium anisopliae* ARSEF 2575 from insect blood  
Metarhizium anisopliae cDNA clone B711 5', mRNA sequence.  
ACCESSION CN808249.1 GI:47729722  
VERSION CN808249.1  
KEYWORDS EST.  
SOURCE Metarhizium anisopliae  
ORGANISM Metarhizium anisopliae  
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
Hypocreomycetidae; Hypocreales; Clavicipitaceae; mitosporic  
Clavicipitaceae; Metarhizium.

REFERENCE 1 (bases 1 to 998)  
AUTHORS Wang, C.S., Hu, G. and St. Leger, R.J.  
TITLE Gene expression profiling of *Metarhizium anisopliae* grown under different conditions: mechanisms of fungal opportunism  
JOURNAL Unpublished (2004)  
COMMENT Contact: Wang CS  
Department of Entomology  
University of Maryland  
4112 Plant Sciences Building, College Park, MD 20742, USA  
Email: cwang4@umd.edu  
Seq primer: M13 Reverse.

FEATURES  
source 1..998  
Location/Qualifiers  
/organism="Metarhizium anisopliae"  
/mol\_type="mRNA"  
/db\_xref="taxon:5530"  
/clone="B711"  
blood\_lib="Metarhizium anisopliae ARSEF 2575 from insect blood"  
/note="Vector: PCMV.SPORT6.1; Metarhizium anisopliae was grown in insect haemolymph for 24 hours. A cDNA library was constructed in the vector PCMV.SPORT6.1"

## ORIGIN

Query Match 9.2%; Score 93.2; DB 7; Length 998;  
Best Local Similarity 49.5%; Pred. No. 1.3e-17;  
Matches 270; Conservative 0; Mismatches 273; Indels 3; Gaps 1;

```
QY 15 CAAGTGCCTGCGTCAAGCCGCCGATCTACATGATTTGGAGCGACGGTGGACAA 74
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 108 CAAGGTCGCTGCGTCAAGCCGCCGATCTACATGATTTGGAGCGACGGTGGACAA 167
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 75 AACATGAGTTGATGAGAAAGACGACAGTAATTAATGCTCGTCTGATGCCCTTCCGA 134
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 168 GACTGTGACTGATCTCGAAGCTGGCGAAGATGTCAAGCTTATGCTTCTCTGA 227
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 135 AACTTGATTCAGGTCACCATGCTTTCTTGGCTTGACTCAGCAGCATGGGCATGCA 194
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 228 AGTTTGATTCCTGATTTCCCTATTGGCTGGCGTGTCACTACCAAGATTTCGCTGCC 287
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 195 ATTTGAGCCCAATACATGAGAACTCATTTGAGTTGATGGCCCTCAAGCTAAGCGCAT 254
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 288 CCTCTCAAGAAAGTTTACACAGAACGATCGGCCGACTGGAGAAATGCGACGAAT 347
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 255 TTCAAGTCAGCCAAAGCGGTGGGAATCATGTGTCACCTTGGGAGTGAAGAAAGGTCGG 314
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 348 TCGCAGGCTGCGCAAGCGGCTAAGATCTACGCTCTGTTGGGATATTCGAGCTGACGG 407
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 315 TGGCACCTTTACATCATGCTAGTGGTTCAATAGCGCATTAATGTGACACCATTTGGGGCCG 374
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 408 CCATACCATGTATCATGCGCAGATCATTCATGATCCACCGGAACCGTATCAACCAACCG 467
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 375 GGGAAAGTTGAAACCTACTTTTGTGAAAGTACTTTGTCGGGGAAGGGGATGGTTCAATC 434
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 468 TCGCAAGATCAAGCCCAACCATGTCGAAAGCTTGTATTTGGCGAGGGAAGCGGTGACTC 527
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 435 GCT---AGCGTTTTCAGACGTCGTGTGAAAGGCTGGGTGGCTTATGCTGTTGGAGCA 491
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 528 TCTTCAAGCGCGTATGAGACTGAATTTGGCAATCTCGGCCACCTGAACTGGGGAATA 587
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 492 CCTTCAACCGCTAACAAATACGCTTTGTATGCACAAAATGAAGATTCATTGGCGGC 551
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 588 TATGAACCCCTTCTCAAGCGCTCAACCGCAGCAAAACGAAAGAAATCAGATAGCCGC 647
   ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

QY 552 TTGGCC 557  
|||||  
Db 648 ATGGCC 653

RESULT 11  
COL37063 515 bp mRNA linear EST 17-JUN-2004  
LOCUS Aspergillus flavus Normalized cDNA Expression Library  
DEFINITION Aspergillus flavus cDNA clone NAFCD28 5' end, mRNA sequence.  
ACCESSION COL37063  
VERSION COL37063.1 GI:48866041  
KEYWORDS EST.  
SOURCE Aspergillus flavus  
ORGANISM Aspergillus flavus  
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Eurotiomycetes;  
Eurotiales; Trichocomaceae; mitosporic Trichocomaceae; Aspergillus.  
REFERENCE Yu,J., Whitelaw,C.A., Nierman,W.C., Bhattacharya,D. and Cleveland,T.E.  
TITLE Aspergillus flavus expressed sequence tags for identification of  
genes with putative roles in aflatoxin contamination of crops  
FEMS Microbiol. Lett. (2004) In press  
CONTACT: Yu J  
Food and Feed Safety Research Unit  
USDA/ARS, Southern Regional Research Center  
1100 Robert R. Lee Boulevard, New Orleans, LA 70124, USA  
Tel: 504 286 4405  
Fax: 504 286 4419  
Email: jinyu@arcc.ars.usda.gov  
Contract Dr. Yu at USDA/ARS SRRC (jinyu@arcc.ars.usda.gov) for clone  
information  
PCR Primers  
FORWARD: M13F  
BACKWARD: M13R  
Seq primer: M13 Forward  
POLYA=No.

FEATURES  
source  
1. 515  
/organism="Aspergillus flavus"  
/mol\_type="mRNA"  
/strain="NRRL 3357"  
/db\_xref="taxon:5059"  
/clone="NAFCD28"  
/sex="asexual mycelia"  
/cell\_type="mycelia"  
/dev\_stage="developmental stages from 18 to 96 hours"  
/lab\_host="E. coli DH10B T1 resistant cells"  
/clone\_lib="Aspergillus flavus Normalized cDNA Expression  
Library"  
/note="Vector: pBluescript (SK+) (Stratagene), antibiotic  
selection marker: Carbenicillin; Site\_1: NotI, at the 5  
prime end; Site\_2: EcoRI, at the 3 prime end; This  
normalized cDNA expression library was constructed using a  
mixture of mycelial cells grown under eight different  
medium conditions and harvested at 5 time points (18, 24,  
48, 72, 96 hours). The poly-A sequence was trimmed off  
before ligating to vector."

ORIGIN  
Query Match 8.7%; Score 88.4; DB 7; Length 515;  
Best Local Similarity 51.0%; Pred. No. 3.6e-16;  
Matches 209; Conservative 0; Mismatches 201; Indels 0; Gaps 0;

QY 14 TCAGGTGCGCTGCTGCAAGCCGCCGATCTACATGATTTGGAGCGAGGTGACA 73  
|||||  
Db 58 TCAGGTGCGCTGCGGTTCAAGCTGAGCTGTCTGAAATGACCTACAAAGCGGGTTTCCA 117  
|||||  
QY 74 AAACCATGATGATGGAAGAGACACGTAATATAGCTGCTGATTCGCTTTCCGG 133  
|||||  
Db 118 AGGTATCTCACTGATCAAGAGCGCTGCAAAAACGCGCCAAAGGTGGTCCGTTCCCG 177  
|||||  
QY 134 AAATTTGATTCAGAGCTAACCCATGTTTCTTTGGCTTGACTCACACAGCATGGGCAATGC 193  
|||||

Db 178 AGGTATTCATCCCTGTTATCCATGAGTATCTGGACTAATCTGGCATTTGAGAAATGTT 237  
|||||  
QY 194 AATTGTACGCCAATATACATAGAAATCATTTGATGGAGCCCTCAAGCTAAGCGCA 253  
|||||  
Db 238 CCTTCATGAACGAGTACTTTCAGAACTCCATGAAAGGAATCGGAGATGGAACGCA 297  
|||||  
QY 254 TTTCAGATGACAGCCAGCGGTTGGGAATCATGTCTACCTCGGGATAGATGACGGTGC 313  
|||||  
Db 298 TCAAAATTTGCCGTGCGGAAGCGGGTGTATTCATCTGTTAGGGTATAGGAACGTTACC 357  
|||||  
QY 314 GTGGACCCCTTTATCATGATGATGAGTTTCATAGGCGAATATAGTGAACCACTTGGGGCCC 373  
|||||  
Db 358 GAGATCATTTGTATATCTGCTAGTCAATTTATCATCTCTACTGATGATCTATGCTCACCAACC 417  
|||||  
QY 374 GCGCAAGTTGAAACCTATTTTGTGAACGTACTTTGTCGCGGAAGG 423  
|||||  
Db 418 GACGGAAGATCAAGCCGATCATGTGAGAAAGCTATTGGGGAGACGGG 467  
|||||

RESULT 12  
CF453420 655 bp mRNA linear EST 01-JAN-2004  
LOCUS fcs36.1 124 Fusarium verticillioides fcc1 mutant subtraction  
DEFINITION library Gibberella moniliformis cDNA, mRNA sequence.  
ACCESSION CF453420  
VERSION CF453420.1 GI:40546865  
KEYWORDS EST.  
SOURCE Gibberella moniliformis  
ORGANISM Gibberella moniliformis  
Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
Hypocreomycetidae; Hypocreales; Nectriaceae; Gibberella.  
REFERENCE Pittilla,A.M., McIntyre,L.M., Payne,G.A. and Woloshuk,C.P.  
TITLE Comparison of gene expression in the wild type and the fcc1 mutant  
of Fusarium verticillioides reveals expressed sequence tags  
associated with fumonisin biosynthesis  
Unpublished (2003)  
JOURNAL  
CONTACT: Woloshuk CP  
Department of Botany and Plant Pathology  
Purdue University  
915W. State Street, West Lafayette IN 47907-2054, USA  
Tel: 765 494 3450  
Fax: 765 494 0363  
Email: woloshuk@purdue.edu  
Insert Length: 655 Std Error: 0.00  
Plate: ft. row: O column: 11  
Seq primer: T7.

FEATURES  
source  
1. 655  
/organism="Gibberella moniliformis"  
/mol\_type="mRNA"  
/strain="FTS36"  
/db\_xref="taxon:117187"  
/clone\_lib="Fusarium verticillioides fcc1 mutant  
subtraction library"  
/note="Vector: pGem-T-Easy; Site\_1: EcoRI, Site\_2: EcoRI;  
Fungus was grown on cracked corn. RNA was isolated using  
phenol LiCl method. PolyA RNA was obtained with Oligotex  
mRNA spin columns (Qiagen). Subtracted from wild type RNA  
with PCR select cDNA subtraction kit (Clontech),  
amplified, cloned into pGem-Tz2 and transformed to E. coli  
DH5 alpha cells."

ORIGIN  
Query Match 8.6%; Score 87; DB 6; Length 655;  
Best Local Similarity 48.6%; Pred. No. 1.1e-15;  
Matches 268; Conservative 0; Mismatches 280; Indels 3; Gaps 1;

QY 15 CAAGTGGCGCTGCTGCAAGCCGCCGATCTACATGATTTGGAGCGAGGTGACA 74  
|||||  
Db 3 CAAGCGCGCTGCTGCTGCACTCTCGAGCTGATGGATGATCTTGAAGCGGGTTCGCA 62  
|||||

```
QY 75 AACATTGATGTGATGAGAAAGACGACGTAATATGCTGATCGCTTTCCGGA 134
    |||||
DB 63 GACCATGACCTTCATCAAGAGAGCCGCGCAAGCTGGCTCAAACTGCTGCCCTTCCCA 122
QY 135 AACTGGATTCAGAGCTACCCATGTTGTTCTTTGGCTTGACTACACGATGGCAATGCA 194
    |||||
DB 123 AGTTGGATCCCTGGATATCTTACTGATGTGAAAGTCACTACCTCAATCCCTCCC 182
QY 195 ATTGAGCCCAATACCATGAGAACTCATGTGAGTTGATGGCCCTCAAGGCAATG 254
    |||||
DB 183 CATGCTGAGGCGCTACCGGAGAACTCCATGCGCGTGACTCTGAGAAATGCGCGTGT 242
QY 255 TTCAATGACGACCAAGCGGTGGGAATCATGTGACCTCGGGAGATGAGAACGGGTGCG 314
    |||||
DB 243 TGTGCGGCGAGCCCGGATACCAAGATCTACGCTCTCCCTCGGCTTCTGAAATTTAGCA 302
QY 315 TGGCAACCTTTACATCATGTCAGTGGTTCTTAAGCGGATTAATGTGTGACACCATTTGGGCGCG 374
    |||||
DB 303 CCGCACTCTTACCTGCGCCCAAGTCCCTCATGAGCCCGATGGCTGGTCAACACCG 362
QY 375 GGAAGATTGAAACCTACTTTTGTGAACTTTGTTGCTGGCGGAAG--GGATGTTG 431
    |||||
DB 363 TGCAGATGCAAGCCCACTCACTGATGAGAGCTTGATACGGAGTGGCTCCGGGGATAC 422
QY 432 ATCGCTAGCGGTTTTCGAGACGCTGTGTGAAAGCTGGGTGGCTTATGCTGTGGAGCA 491
    |||||
DB 423 CTTCATGGCGGCTTAGCGAGACTGACATCGCGCTGTGGCCAGCTTAACTGCTGGAGAA 482
QY 492 CTTCAACCGCTAACAAATATGCGCTTTGTATGCAAAAATGAAGATTTGTTGCGCG 551
    |||||
DB 483 CATGAACCCCTTCTCAAGTCTCTCAACGTTCTGTGATGAGAGGTTCAATCGCTGC 542
QY 552 TTGGCCGAGCT 562
    |||||
DB 543 TTGGCCCGTCT 553
    |||||
```

```
RESULT 13
DR631467 699 bp mRNA linear EST 11-JUL-2005
LOCUS EST1021595 Fv1 Gibberella moniliformis cDNA clone FVIE956, mRNA
DEFINITION sequence.
ACCESSION DR631467
VERSION DR631467.1 GI:70706307
KEYWORDS EST.
SOURCE Gibberella moniliformis
ORGANISM Gibberella moniliformis
REFERENCE Gibberella moniliformis
AUTHORS Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
Hypocrosmiidae; Hypocreales; Nectriaceae; Gibberella.
1 (bases 1 to 699)
Brown, D.W., Cheung, F., Proctor, R.H., Butchko, A.E., Zheng, L., Lee, Y.,
Utecherick, T., Smith, S., Feldblyum, T., Glenn, A.B., Plattner, R.D.,
Kendra, D.F., Town, C.D. and Whitehead, C.A.
Analysis of 87,000 expressed sequence tags reveals alternatively
spliced introns in multiple genes of the fumonisin gene cluster
Unpublished (2005)
JOURNAL Contact: Brown, D.W.
COMMENT USDA/ARS/NCADR
USDA
1815 N. University St, Peoria, IL 61604, USA
Tel: 309 681 6230
Fax: 309 681 6689
Email: brown@ncaur.usda.gov
RIGR sequence name: FVIE956TH
Seq primer: AAT TAA CCC TCA AAG GG.
Location/Qualifiers
```

## FEATURES

```
source
1..699
/organism="Gibberella moniliformis"
/mol_type="mRNA"
/strain="m3125"
/db_xref="taxon:117187"
/clone="FVIE956"
/issue_type="mycelia"
```

```
/clone.lib="Fv1"
/note="Vector: pBluescript II SK(+); XR: Site 1: EcoRI;
Site 2: XhoI; anamorph: Fusarium verticillioides. Library
Fv1 was prepared from growth on excised maize seedling
roots and shoots. The roots and shoots were prepared as
described above and then inoculated by dipping briefly in
a suspension of 5 x 106 conidia per ml. The inoculated
roots and shoots were then incubated on moistened Whatman
#1 filter paper at room temperature under sterile
conditions for 4 days. Fungal mycelial was collected,
frozen in liquid nitrogen, ground to a powder, and then
added to TRIzol Reagent (Invitrogen, Carlsbad CA) at
approximately 1 g mycelia per 10 ml TRIzol. The cDNA was
directionally ligated into the pBluescript II SK(+) XR
vector (cDNA Synthesis Kit; Stratagene)."
```

## ORIGIN

```
Query Match 8.6%; Score 87; DB 8; Length 699;
Best Local Similarity 48.6%; Pred. No. 1.1e-15;
Matches 268; Conservative 0; Mismatches 280; Indels 3; Gaps 1;
```

```
QY 15 CAAGTGCCTGCGTGCAGCGCCCGATCTACATGATTTGAGGCGAGGTGACAA 74
    |||||
DB 68 CAAGGCGCGCTGCTGTCACCTCCAGCTGATGTTGATGAGGCGGTTGCGAA 127
QY 75 AACATTGATGTGATGAGAAAGACGACGTAATATGCTGCTGATTCGCTTTCCGGA 134
    |||||
DB 128 GACCATGACTTCATCAACGAGCGCGCAAGCTGGCTGCAAACTGTGCGCTTCCGA 187
QY 135 AACTGAGATTCAGAGCTACCATGTTCTTTGCTTACTCACCGAGATGGCAATGCA 194
    |||||
DB 188 AGTTGGATCCCTGGATATCTTACTGATGTGAAAGTCACTACCTCCATCCCTCC 247
QY 195 ATTGTACGCCAATACATGAGAACTCATTTGAGTTGANTGAGCCCTCAAGCTTAAGCGCAT 254
    |||||
DB 248 CATGCTGAGGCGCTACCGGAGAACTCCATGAGCGCTGCACTGAGAAATGCGCGTAT 307
QY 255 TTCAGATGACCGCAAGCGGTGGAGATCATGCTCACTCGGGAGATGAGTAAAGGTGCG 314
    |||||
DB 308 TGTGCGCGACCGCGCATTAACAGATCTACGCTCTCCCTGCTTCTGTGAAATTTAGCA 367
QY 315 TGGCAACCTTTACATCATGCTGATGAGTGAATGAGTATGATGACACATTTGGGCGCG 374
    |||||
DB 368 CGCAACTCTTACCTCGCGCAAGTCTCATGCGCCGATGGCTCGGTCAACACACCG 427
QY 375 GCGAAAGTTGAAACCTACTTTGTTGAACTGTTGTTGTTGCGGAAAG--GGATGTTG 431
    |||||
DB 428 TCGCAAGATCAAGCACTCACTGTTGAGAACCTTGATTAACGCGATGGCTCCGGGATATC 487
QY 432 ATCGCTAGCGGTTTTCGAGACGCTGTGTGAAAGCTGGGTGGCTTATGCTGTGGAGCA 491
    |||||
DB 488 CTTCATGGCGGCTTAGCGAGATGATCGCGCGTGTGGCCAGCTTAACTCTGGAGAA 547
QY 492 CTTCAACCGGTTACAAATATGCGCTTTGTATGCAAAATGAGATTCATTGCGCGG 551
    |||||
DB 548 CATGAACCCCTTCTCAAGTCTCTCAACGTTTCTGTGATGAGAGGTTCAATCGCTGC 607
QY 552 TTGGCCGAGCT 562
    |||||
DB 608 TTGGCCCGTCT 618
    |||||
```

```
RESULT 14
DR631867 732 bp mRNA linear EST 11-JUL-2005
LOCUS EST1021995 Fv1 Gibberella moniliformis cDNA clone FVIE80, mRNA
DEFINITION sequence.
ACCESSION DR631867
VERSION DR631867.1 GI:70706707
KEYWORDS EST.
SOURCE Gibberella moniliformis
ORGANISM Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;
```

REFERENCE 1 (bases 1 to 732) Hypocreomycetidae; Hypocreales; Nectriaceae; Gibberella.  
 AUTHORS Brown,D.W., Cheung,F., Proctor,R.H., Butchko,A.E., Zheng,L., Lee,Y., Utechtack,T., Smith,S., Feldblyum,T., Glenn,A.E., Plattner,R.D., Kendra,D.F., Town,C.D. and Whitelaw,C.A.  
 TITLE Analysis of 87,000 expressed sequence tags reveals alternatively spliced introns in multiple genes of the fumonisin gene cluster unpublished (2005)  
 JOURNAL Contact: Brown, D.W.  
 COMMENT USDA/ARS/NCAR  
 USDA 1815 N. University St, Peoria, IL 61604, USA  
 Tel: 309 681 6230  
 Fax: 309 681 6689  
 Email: brown@ncar.usda.gov  
 TIGR sequence name: FVIEB80TH  
 Seq primer: AAT TAA CCC TCA AAG GG.  
 FEATURES  
 source 1..732  
 /organism="Gibberella moniliformis"  
 /mol\_type="mRNA"  
 /strain="m3125"  
 /db\_xref="taxon:117187"  
 /clone="FVIEB80"  
 /tissue\_type="mycelia"  
 /clone\_lib="FVI"  
 /note="Vector: pBluescript II SK(+) XR; Site 1: EcoRI; Site 2: XhoI; anamorph: Fusarium verticillioides. Library FVI was prepared from growth on excised maize seedling roots and shoots. The roots and shoots were prepared as described above and then inoculated by dipping briefly in a suspension of 5 x 10<sup>6</sup> conidia per ml. The inoculated roots and shoots were then incubated on moistened Whatman #1 filter paper at room temperature under sterile conditions for 4 days. Fungal mycelial was collected, frozen in liquid nitrogen, ground to a powder, and then added to Trizol Reagent (Invitrogen, Carlsbad CA) at approximately 1 g mycelia per 10 ml Trizol. The cDNA was directionally ligated into the pBluescript II SK(+) XR vector (CDNA Synthesis Kit; Stratagene)."

ORIGIN  
 Query Match 8 6%; Score 87; DB 8; Length 732;  
 Best Local Similarity 48.6%; Pred. No. 1.1e-15;  
 Matches 268; Conservative 0; Mismatches 280; Indels 3; Gaps 1;

15 CAAGTGCCTGCTGCAAGCCGCCGATCTACATGATTGGAGCGACGGTGGACAA 74  
 |||||  
 68 CAAGGCGCTGCTGCAACCTCGAGCTGGATGCTTGAAGGCGGTGTCGCAA 127  
 |||||  
 75 AACCATTTGATTGAGAGAGACAGACGTAATAGCTGCTGATGGCTTTCGGA 134  
 |||||  
 128 GACCATGACCTTATCAACAGAGCGCGCAAGCTGCTGCAACTGTCCTTTCGA 187  
 |||||  
 135 AACCTGGATTCCAGGCTACCATGCTTCTTGGCTTGACTACACAGCATGGGCAATGA 194  
 |||||  
 188 AGTTTGATCCCTGGAATCTTACTGATGTGAAGGTCACTACCTCCATCCCTCCC 247  
 |||||  
 195 ATTGTAGCCCAATACCATGAGAACTCATTTGAGATTGAGCCCTCAAGTGAAGCAT 254  
 |||||  
 248 CATGCTTAACCGCTACCGCGAATCTCATGGCGCTGACTGAGGAATATGCGCGTAT 307  
 |||||  
 255 TTGAGATGCAAGCGGTTGGGAATCATGTCACCTGGGATGAGTGAACGGGTCGG 314  
 |||||  
 308 TCGTGGCGAGCCCGCATTAACGATCTACGTCCTCCCTGCGCTTCTCTAGATTGACCA 367  
 |||||  
 315 TGGACCCCTTTACATCAGTCAAGTGTTCATAGGCGAATATGTGACCACTTGGGCGCG 374  
 |||||  
 368 CGCAACTCTTACTCGCCCAAGTCTCATGCGCCCGAATGGCTCGGTCAACAACACCG 427  
 |||||  
 375 GCGAAGTGAACCTTACTTTTGTGAACGACTTTGTGGCGGAAG---GATGGTTC 431  
 |||||  
 428 TCGCAAGATCAAGCAACTCACTGTGAAGACTGTATACGGCGATGCTCCGGGGATAC 487  
 |||||

QY 432 ATCGCTAGCGTTTTCGAGACGCTCTTTGGAAGGCTGGTGGCTTANCGTTGGAGCA 491  
 |||||  
 DB 488 CTTCAATGGCGTTTACGAGACATCGGCGGTGTGGCCAGCTTAACCTGTGGAGAA 547  
 |||||  
 QY 492 CTTTCAACGGCTACAGAAATATGCGTTTGTATGACAAATGAGATTCATTGGCGCG 551  
 |||||  
 DB 548 CATGAACCCCTTCCTCAACTCTCTCAAGCTTCTGTGTGTGAGCAGGTTCAATGCTGC 607  
 |||||  
 QY 552 TTGGCCGAGCT 562  
 |||||  
 DB 608 TTGGCCGCTCT 618  
 |||||

RESULT 15  
 DR629486 810 bp mRNA linear EST 11-JUL-2005  
 LOCUS EST1019614 FVI Gibberella moniliformis cDNA clone FVIDC93, mRNA  
 DEFINITION  
 sequence.  
 DR629486  
 DR629486 GI:70704200  
 VERSION  
 EST.  
 SOURCE  
 KEYWORDS  
 ORGANISM  
 Gibberella moniliformis  
 Eukaryota; Fungi; Ascomycota; Pezizomycotina; Sordariomycetes;  
 Hypocreomycetidae; Hypocreales; Nectriaceae; Gibberella.  
 REFERENCE 1 (bases 1 to 810)  
 Brown,D.W., Cheung,F., Proctor,R.H., Butchko,A.E., Zheng,L., Lee,Y., Utechtack,T., Smith,S., Feldblyum,T., Glenn,A.E., Plattner,R.D., Kendra,D.F., Town,C.D. and Whitelaw,C.A.  
 TITLE Analysis of 87,000 expressed sequence tags reveals alternatively spliced introns in multiple genes of the fumonisin gene cluster unpublished (2005)  
 JOURNAL Contact: Brown, D.W.  
 COMMENT USDA/ARS/NCAR  
 USDA 1815 N. University St, Peoria, IL 61604, USA  
 Tel: 309 681 6230  
 Fax: 309 681 6689  
 Email: brown@ncar.usda.gov  
 TIGR sequence name: FVIDC93TH  
 Seq primer: AAT TAA CCC TCA AAG GG.  
 FEATURES  
 source 1..810  
 /organism="Gibberella moniliformis"  
 /mol\_type="mRNA"  
 /strain="m3125"  
 /db\_xref="taxon:117187"  
 /clone="FVIDC93"  
 /tissue\_type="mycelia"  
 /clone\_lib="FVI"  
 /note="Vector: pBluescript II SK(+) XR; Site 1: EcoRI; Site 2: XhoI; anamorph: Fusarium verticillioides. Library FVI was prepared from growth on excised maize seedling roots and shoots. The roots and shoots were prepared as described above and then inoculated by dipping briefly in a suspension of 5 x 10<sup>6</sup> conidia per ml. The inoculated roots and shoots were then incubated on moistened Whatman #1 filter paper at room temperature under sterile conditions for 4 days. Fungal mycelial was collected, frozen in liquid nitrogen, ground to a powder, and then added to Trizol Reagent (Invitrogen, Carlsbad CA) at approximately 1 g mycelia per 10 ml Trizol. The cDNA was directionally ligated into the pBluescript II SK(+) XR vector (CDNA Synthesis Kit; Stratagene)."

ORIGIN  
 Query Match 8 6%; Score 87; DB 8; Length 810;  
 Best Local Similarity 48.6%; Pred. No. 1.2e-15;  
 Matches 268; Conservative 0; Mismatches 280; Indels 3; Gaps 1;

15 CAAGTGCCTGCTGCAAGCCGCCGATCTACATGATTGGAGCGACGGTGGACAA 74  
 |||||



```

Db      67  CAAAGCCGCTGCTCTCACTCCAGCCTGATGGTTGATCTTGAGGGCGGTGTTCCAA 126
QY      75  AACCATGATGTTGATGGAAGAAGCAGACGTAATATGCTGTGATTCGCTTTCCGA 134
Db      127  GACCATGATCTTCACTCAAGAGCCGGCCAACTGGCTGCAAACTCGTCCCTTCCGA 186
QY      135  AACTTGATTCAGGCTAACCCATGGTTTCTTGGCTTGACTCACACAGCATGGCAATGCA 194
Db      187  AGTTTGATCCCTGATATCCTTACTGGATGTGGAAGTCACTAATCCATCCCTCCC 246
QY      195  ATTGTACGCCAATACATGAGAACTCATTTGAGTTGATGGCCCTCAAGCTAAGGCAT 254
Db      247  CATGCTGAAGCGCTACCGCAACTCCATGGCCGCTGACTTGAGGAAATGCGCCGTAT 306
QY      255  TTCGATGACAGCCCAAGCGGTTGGGAATCATGGTCAACCCTGGGGATGAGTGAACGGTCCG 314
Db      307  TCGTCGCGCAGCCCGCATTAACAGATCTACGTCTCCCTCGGCTTCTGTGAGATTGACA 366
QY      315  TGGCACCCCTTACATCAGTAGTGTTCATAGCGCATATGATGACACATTTGGGGCCCG 374
Db      367  CGCAACTCTTTACCTGGCCCAAGTCTCATCGGCCCGATGGCTCGGTCAATCAACGACCG 426
QY      375  GCGAAAGTTGAAACTTACTTTGTTGAACGTACTTTGTCGGCGAAGGGGATGG--TTC 431
Db      427  TCGCAAGATCAAGCACTCACGTTGGAAGCTTGTATACGGCGATGGCTCCGGGGATAC 486
QY      432  ATGCTAGCGGTTTGGAGAAGTCTGTGGAAGGCTGGGCTTATGCTGTGGAGCA 491
Db      487  CTTCATGGCCGTTAGCGAGACTGACATCGGCCGTTGGCCAGCTTAACTGCTGGAGAA 546
QY      492  CCTTCAACCGCTTAACAATAATCGCTTGTATGCACAAAATGAAGATTCATTGTGGCGC 551
Db      547  CATGAACCCCTTCTCTCAAGTCTCTCAACGTTTCTGCTGTGAGCAGGTTCAATCGCTGC 606
QY      552  TTGGCCGAGCT 562
Db      607  TTGGCCGCTCT 617

```

Search completed: April 27, 2006, 01:35:53  
 Job time : 4780.42 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 27, 2006, 00:01:50 ; Search time 5473.13 Seconds  
(without alignments)  
10531.310 Million cell updates/sec

Title: US-09-751-299-3  
Perfect score: 1014  
Sequence: 1 atgaagaagatcatcaagt.....cgcaactcgaagaatttga 1014

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 5883141 seqs, 28421725653 residues  
Total number of hits satisfying chosen parameters: 11766282

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl:\*  
1: gb\_ba:\*  
2: gb\_in:\*  
3: gb\_env:\*  
4: gb\_om:\*  
5: gb\_ov:\*  
6: gb\_pat:\*  
7: gb\_ph:\*  
8: gb\_pr:\*  
9: gb\_ro:\*  
10: gb\_sts:\*  
11: gb\_sy:\*  
12: gb\_un:\*  
13: gb\_vl:\*  
14: gb\_htg:\*  
15: gb\_pl:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query	Length	DB ID	Description
1	1014	100.0	1014	3	AY487562 Unculture
2	1014	100.0	1014	6	AX189646 Sequence
3	499.4	49.3	1014	3	AY487445 Unculture
4	497.8	49.1	1014	3	AY487449 Unculture
5	427.4	42.1	110000	1	CP000075 Pseudomon
6	412.8	40.7	110000	1	AE016853_02 Continuation (3 of
7	399.6	39.4	999	3	AY487512 Unculture
8	396.4	39.1	1002	3	AY487492 Unculture
9	364.2	35.9	1038	3	AY487427 Unculture
10	319	31.5	1011	3	AY487546 Unculture
11	311.4	30.7	1026	3	AY487438 Unculture
12	309.4	30.5	1077	3	AY487522 Unculture
13	308.4	30.4	1047	3	AY487430 Unculture
14	307	30.3	1062	3	AY487543 Unculture
15	305.4	30.1	1041	3	AY487497 Unculture
16	305.4	30.1	1041	6	AX189644 Sequence
17	304.8	30.1	1065	3	AY487496 Unculture
18	300.4	29.6	1062	3	AY487552 Unculture

19	294	29.0	110000	1	AP006618_34 Continuation (35 o
20	289.2	28.5	1050	3	AY487432 Unculture
21	289	28.5	1014	3	AY487464 Unculture
22	288	28.4	1014	3	AY487429 Unculture
23	287.8	28.4	6976	1	AY885240 Pseudomon
24	286.2	28.2	1059	6	CQ874190 Sequence
25	284.6	28.1	1011	3	AY487507 Unculture
26	283.4	27.9	1017	3	AY487558 Unculture
27	283	27.9	1017	3	AY487475 Unculture
28	283	27.9	1035	6	AR036234 Sequence
29	283	27.9	1200	6	AR036235 Sequence
30	283	27.9	1200	6	E12616 DNA encodin
31	281.4	27.8	1017	3	AY487536 Unculture
32	281	27.7	1014	3	AY487463 Unculture
33	280.2	27.6	1005	3	AY487559 Unculture
34	279.6	27.6	105027	1	AE17740 Pseudomon
35	278.4	27.5	1014	3	AY487515 Unculture
36	276.6	27.3	1014	3	AY487473 Unculture
37	274.4	27.1	984	3	AY487523 Unculture
38	273	26.9	1017	3	AY487436 Unculture
39	271.4	26.8	1017	3	AY487480 Unculture
40	269.8	26.6	1017	3	AY487531 Unculture
41	263.4	26.0	1017	3	AY487448 Unculture
42	261	25.7	110000	1	BA000040_70 Continuation (71 o
43	258.8	25.5	1008	3	AY487545 Unculture
44	257	25.3	1041	3	AY487481 Unculture
45	256	25.2	1059	3	AY487532 Unculture

## ALIGNMENTS

RESULT 1	AY487562	1014 bp	DNA	linear	ENV 05-APR-2004
LOCUS	AY487562	Uncultured organism clone 2A6 nitrilase (BD5220) gene, complete cds.			
DEFINITION	AY487562.1	GI:40890328			
ACCESSION	AY487562	ENV.			
VERSION	AY487562.1	GI:40890328			
KEYWORDS	ENV.	uncultured organism			
SOURCE	ENV.	uncultured organism			
ORGANISM	ENV.	uncultured organism			
REFERENCE	1 (bases 1 to 1014)	unclassified; environmental samples.			
AUTHORS	Robertson,D.E., Chaplin,J.A., Desantis,G., Podar,M., Madden,M., Chi,E., Richardson,T.H., Milan,A., Miller,M., Weiner,D.P., Wong,K., McQuaid,J., Farwell,B., Preston,L.A., Tan,X., Snead,M.A., Keller,M., Mathur,E., Kretz,P.L., Burk,M.J. and Short,J.M.				
TITLE	Exploring nitrilase sequence space for enantioselective catalysis Appl. Environ. Microbiol. 70 (4), 2429-2436 (2004)				
JOURNAL	2 (bases 1 to 1014)				
REFERENCE	Robertson,D.E., Chaplin,J.A., Desantis,G., Podar,M., Madden,M., Chi,E., Richardson,T.H., Milan,A., Miller,M., Weiner,D.P., Wong,K., McQuaid,J., Farwell,B., Preston,L.A., Tan,X., Keller,M., Mathur,E., Kretz,P.L., Burk,M.J. and Short,J.M.				
AUTHORS	Direct Submission				
TITLE	Submitted (26-NOV-2003) Bioinformatics, Diversa Corporation, 4955				
JOURNAL	Directors Place, San Diego, CA 92121, USA				
FEATURES	Location/Qualifiers				
source	1..1014				
	/organism="uncultured organism"				
	/mol_type="genomic DNA"				
	/db_xref="taxon:155900"				
	/clone="2A6"				
	/environmental_sample				
gene	1..1014				
	/locus_tag="BD5220"				
CDS	1..1014				
	/locus_tag="BD5220"				
	/EC_number="3.5.5.7"				
	/codon_start=1				
	/evidence=experimental				
	/product="nitrilase"				

/protein\_id="AAR97509.1"  
/db\_xref="GI:40890329"  
/translation="MKRAIVKACVOAIPYMDLEATVDKTIEMBEAARNNAFLIAP  
ETWIPGYPWFLMIDSPAMAMQFVROYHENSELEIDGPAKRIISAARKIGIMTLGMBE  
RVGGTLYISQMFIDNGDITIGARKKLPTEVERTLREBGSSLIAPETISVGLGIC  
CWEHLPLTYALYQNEIEHCAWPSFSLYPNAAKLGDVVAASRIYAVBQCFTV  
LASCALVSOSIMDLCTDDEKHALLLAGGSHRIIGDGDVLVAPLAENEGLIVANL  
DPGVRIILAKMAADPAHYSRPDITRLIDRSPKLPVEIEBGDIRPYALGKASBTGAOL  
EET"

ORIGIN

Query Match 100.0%; Score 1014; DB 3; Length 1014;  
Best Local Similarity 100.0%; Pred. No. 1.1e-290; Mismatches 0; Gaps 0;  
Matches 1014; Conservative 0; Indels 0; Gaps 0;  
QY 1 ATGAAAAGCTATCAAGGTGCGCTGGTGCAACCGCCCGATCTCATGAGATTGGAG 60  
DB 1 ATGAAAAGCTATCAAGGTGCGCTGGTGCAACCGCCCGATCTCATGAGATTGGAG 60  
QY 61 GCGACGGTGACAAACCATTTGATGAGAGAGACGCTAAATATGCTGCTG 120  
DB 61 GCGACGGTGACAAACCATTTGATGAGAGAGACGCTAAATATGCTGCTG 120  
QY 121 ATCGCCTTTCCGGAACCTTGAGTTCCAGGCTACCCATGCTTTGGCTTGACTCA 180  
DB 121 ATCGCCTTTCCGGAACCTTGAGTTCCAGGCTACCCATGCTTTGGCTTGACTCA 180  
QY 181 GCATGGGCAATGCAATTTGTACGCCAATACCATGAGAACTCATTTGAGTGGCCCT 240  
DB 181 GCATGGGCAATGCAATTTGTACGCCAATACCATGAGAACTCATTTGAGTGGCCCT 240  
QY 241 CAAGCTAAGCGCATTTCAGATGACGACCAAGCGGTGGGAATCATGCTGACCTGGGGATG 300  
DB 241 CAAGCTAAGCGCATTTCAGATGACGACCAAGCGGTGGGAATCATGCTGACCTGGGGATG 300  
QY 301 AGTGAACGGGTGCGTGAGACCCCTTACATCACTAGTGGTTCATAGGGGATTAATGGTAC 360  
DB 301 AGTGAACGGGTGCGTGAGACCCCTTACATCACTAGTGGTTCATAGGGGATTAATGGTAC 360  
QY 361 ACCATTGGGGCCCGGCGAAAGTTGAACCTTATTTGTTGAACGTACTTTGTTGGCGAA 420  
DB 361 ACCATTGGGGCCCGGCGAAAGTTGAACCTTATTTGTTGAACGTACTTTGTTGGCGAA 420  
QY 421 GGGGATGCTTCATGCTGACGCGTTTTCAGACGCTGTTGGAAGGCTGGGTGGCTTAAGC 480  
DB 421 GGGGATGCTTCATGCTGACGCGTTTTCAGACGCTGTTGGAAGGCTGGGTGGCTTAAGC 480  
QY 481 TGTGGGAGACCTTCAACCGCTACAAATATGCGCTTGTATGCAAAATGAAAGATT 540  
DB 481 TGTGGGAGACCTTCAACCGCTACAAATATGCGCTTGTATGCAAAATGAAAGATT 540  
QY 541 CATTTGGGGCTTGGCCGAGCTTTAGCCTTATCTTAATGCGGGAAGCCCTGGGGCCT 600  
DB 541 CATTTGGGGCTTGGCCGAGCTTTAGCCTTATCTTAATGCGGGAAGCCCTGGGGCCT 600  
QY 601 GATGTCATGTAAGCGGCTCTGCAATCTATGCGGTTGAAGGGCAATGCTCTGACTAGG 660  
DB 601 GATGTCATGTAAGCGGCTCTGCAATCTATGCGGTTGAAGGGCAATGCTCTGACTAGG 660  
QY 661 TCGTGGGCGCTGCTTCAATTCATGATGATGCTTTGTACAGATGACGAAAAGCAT 720  
DB 661 TCGTGGGCGCTGCTTCAATTCATGATGATGCTTTGTACAGATGACGAAAAGCAT 720  
QY 721 GCGTTGCTTCTGGCTGGTGGAGACCTACGATATAGGGCTGATGGTGTGA 780  
DB 721 GCGTTGCTTCTGGCTGGTGGAGACCTACGATATAGGGCTGATGGTGTGA 780  
QY 781 GTGGGCGCTTCTGGCGAAATGAAAGGGTATTTCTTACGCAAACTTTGATCCTGAGAT 840  
DB 781 GTGGGCGCTTCTGGCGAAATGAAAGGGTATTTCTTACGCAAACTTTGATCCTGAGAT 840  
QY 841 GCGATCCTTGGCTAAATGGCGGAGACCTGCTGGCTATTTCCCGTCCGACATTA 900  
DB 841 GCGATCCTTGGCTAAATGGCGGAGACCTGCTGGCTATTTCCCGTCCGACATTA 900

DB 841 GCGATCCTTGGCTAAATGGCGGAGACCTGCTGGCTATTTCCCGTCCGACATTA 900  
QY 901 CGCTTGCTAATGATGACGCCCTTAATTAACGGGTGTTGAAATGAAAGTATCTTGGT 960  
DB 901 CGCTTGCTAATGATGACGCCCTTAATTAACGGGTGTTGAAATGAAAGTATCTTGGT 960  
QY 961 CCTTACGCTTTGGGTAAAGCTGTGAGCGGGTGGCACTCGAAGAAATTTGA 1014  
DB 961 CCTTACGCTTTGGGTAAAGCTGTGAGCGGGTGGCACTCGAAGAAATTTGA 1014

RESULT 2

AX189646 1014 bp DNA linear PAT 08-AUG-2001  
LOCUS AX189646  
DEFINITION Sequence 3 from Patent WO0148175.  
ACCESSION AX189646  
VERSION AX189646.1 GI:15143036  
KEYWORDS  
SOURCE  
ORGANISM  
REFERENCE  
1 Madsen,M.D. and Chaplin,J.A.  
AUTHORS  
TITLE  
METHODS for producing enantiomerically pure \_g(a)-substituted  
carboxylic acids  
JOURNAL  
PATENT: WO 0148175-A 3 05-JUL-2001;  
DIVERSA CORPORATION (US)  
FEATURES  
source  
Location/Qualifiers  
1. 1014  
/organism="unidentified"  
/mol\_type="unassigned DNA"  
/db\_xref="taxon:32644"  
/note="Description of Unknown Organism: Obtained from an  
environmental s ampl"  
1. 1014  
/note="unnamed protein product"  
/codon\_start=1  
/protein\_id="CAC50777.1"  
/db\_xref="GI:15143037"  
/translation="MKRAIVKACVOAIPYMDLEATVDKTIEMBEAARNNAFLIAP  
ETWIPGYPWFLMIDSPAMAMQFVROYHENSELEIDGPAKRIISAARKIGIMTLGMBE  
RVGGTLYISQMFIDNGDITIGARKKLPTEVERTLREBGSSLIAPETISVGLGIC  
CWEHLPLTYALYQNEIEHCAWPSFSLYPNAAKLGDVVAASRIYAVBQCFTV  
LASCALVSOSIMDLCTDDEKHALLLAGGSHRIIGDGDVLVAPLAENEGLIVANL  
DPGVRIILAKMAADPAHYSRPDITRLIDRSPKLPVEIEBGDIRPYALGKASBTGAOL  
EET"

CDS

QY 1 ATGAAAAGCTATCAAGGTGCGCTGGTGCAACCGCCCGATCTCATGAGATTGGAG 60  
DB 1 ATGAAAAGCTATCAAGGTGCGCTGGTGCAACCGCCCGATCTCATGAGATTGGAG 60  
QY 61 GCGACGGTGACAAACCATTTGATGAGAGAGACGCTAAATATGCTGCTG 120  
DB 61 GCGACGGTGACAAACCATTTGATGAGAGAGACGCTAAATATGCTGCTG 120  
QY 121 ATCGCCTTTCCGGAACCTTGAGTTCCAGGCTACCCATGCTTTGGCTTGACTCA 180  
DB 121 ATCGCCTTTCCGGAACCTTGAGTTCCAGGCTACCCATGCTTTGGCTTGACTCA 180  
QY 181 GCATGGGCAATGCAATTTGTACGCCAATACCATGAGAACTCATTTGAGTGGCCCT 240  
DB 181 GCATGGGCAATGCAATTTGTACGCCAATACCATGAGAACTCATTTGAGTGGCCCT 240  
QY 241 CAAGCTAAGCGCATTTCAGATGACGACCAAGCGGTGGGAATCATGCTGACCTGGGGATG 300  
DB 241 CAAGCTAAGCGCATTTCAGATGACGACCAAGCGGTGGGAATCATGCTGACCTGGGGATG 300  
QY 301 AGTGAACGGGTGCGTGAGACCCCTTACATCACTAGTGGTTCATAGGGGATTAATGGTAC 360  
DB 301 AGTGAACGGGTGCGTGAGACCCCTTACATCACTAGTGGTTCATAGGGGATTAATGGTAC 360

ORIGIN

Query Match 100.0%; Score 1014; DB 6; Length 1014;  
Best Local Similarity 100.0%; Pred. No. 1.1e-290; Mismatches 0; Gaps 0;  
Matches 1014; Conservative 0; Indels 0; Gaps 0;  
QY 1 ATGAAAAGCTATCAAGGTGCGCTGGTGCAACCGCCCGATCTCATGAGATTGGAG 60  
DB 1 ATGAAAAGCTATCAAGGTGCGCTGGTGCAACCGCCCGATCTCATGAGATTGGAG 60  
QY 61 GCGACGGTGACAAACCATTTGATGAGAGAGACGCTAAATATGCTGCTG 120  
DB 61 GCGACGGTGACAAACCATTTGATGAGAGAGACGCTAAATATGCTGCTG 120  
QY 121 ATCGCCTTTCCGGAACCTTGAGTTCCAGGCTACCCATGCTTTGGCTTGACTCA 180  
DB 121 ATCGCCTTTCCGGAACCTTGAGTTCCAGGCTACCCATGCTTTGGCTTGACTCA 180  
QY 181 GCATGGGCAATGCAATTTGTACGCCAATACCATGAGAACTCATTTGAGTGGCCCT 240  
DB 181 GCATGGGCAATGCAATTTGTACGCCAATACCATGAGAACTCATTTGAGTGGCCCT 240  
QY 241 CAAGCTAAGCGCATTTCAGATGACGACCAAGCGGTGGGAATCATGCTGACCTGGGGATG 300  
DB 241 CAAGCTAAGCGCATTTCAGATGACGACCAAGCGGTGGGAATCATGCTGACCTGGGGATG 300  
QY 301 AGTGAACGGGTGCGTGAGACCCCTTACATCACTAGTGGTTCATAGGGGATTAATGGTAC 360  
DB 301 AGTGAACGGGTGCGTGAGACCCCTTACATCACTAGTGGTTCATAGGGGATTAATGGTAC 360



Db	602	AAGTAAATGTGCGAGCTTCTGGATCTACGCCCGTGAAGGCAAGTGTGTTGTCGGCTT	661
QY	662	CGTGTGCGCTCGTTCAACATCCATGATCGATATGCTTTGTACAGATGACGAAAGCATG	721
Db	662	CTTCGCGGCTGTCTCGCCAGAGATGATCGAAATGCTCTGACGACGAAAGCAGACAA	721
QY	722	CGTGTGCTTGTGGCTGTGTGTGACACATCAGCTATCATGAGGCGCTGATGATGATGTG	781
Db	722	GCTTCTTTACAGCCCGGCGCGGGTACTCCGCATTATCGGTCCGATGGACGACCTAG	781
QY	782	TCGCGCCTCTTCCGAAATGAAAGGATTTCTCTACGAAACCTTGATCTTGAATAC	841
Db	782	CGCCCCCTTGGGGGAAACAGAGAGGATATCTCTATGCCACTGTGACCTCCGCTG	841
QY	842	GCATTCCTTGCTAAATATGGCGGAGACCTGTGCTGTCATTATTTCCGATCCGACATTA	901
Db	842	GAATCTATGCAAAAGACCGAGTGTATCCAGCCGSCACTTATCCAGACCAACGTCACTC	901
QY	902	GCTTGCTATATAGATGACGACCCCTAAATTACCGGTAGTTGAAATTGAAGTGAATCTTG	961
Db	902	GGCTGCTGATCAATGCGAGTGCATTCAGCCAGTGTGTAGAGTTGGAAGGGAATATCTG	961
QY	962	CTTACGCTTTGGGTAAAGGTGTGAGAGGGGTGGCACACTGAAAGAAATTGA	1014
Db	962	CATCGGCCAAGGCTTTGATGATGGCGGGCCCCCGGGTACGAAGGCAATTGA	1014

RESULT 4	1014 bp	DNA	linear	ENV 05-APR-2004
AY487449	1014 bp	DNA	linear	ENV 05-APR-2004
AY487449	1014 bp	DNA	linear	ENV 05-APR-2004
LOCUS	1014 bp	DNA	linear	ENV 05-APR-2004
DEFINITION	Uncultured organism clone 2A7 nitrilase (BD5275) gene, complete cds.			
ACCESSION	AY487449			
VERSION	AY487449.1			
KEYWORDS	ENV.			
SOURCE	uncultured organism			
ORGANISM	uncultured organism			
REFERENCE	unclassified: environmental samples.			
AUTHORS	1 (bases 1 to 1014)			
TITLE	Robertson, D.E., Chaplin, J.A., Desantis, G., Podar, M., Madden, M., Chi, E., Richardson, T.H., Milán, A., Miller, M., Weiner, D.P., Wong, K., McQuaid, J., Farrell, B., Preston, L.A., Tan, X., Sneed, M.A., Keller, M., Mathur, E., Kretz, P.L., Burk, M.J., and Short, J.M.			
JOURNAL	Exploring nitrilase sequence space for enantioselective catalysis Appl. Environ. Microbiol. 70 (4), 2429-2436 (2004)			
REFERENCE	2 (bases 1 to 1014)			
AUTHORS	Robertson, D.E., Chaplin, J.A., Desantis, G., Podar, M., Madden, M., Chi, E., Richardson, T.H., Milán, A., Miller, M., Weiner, D.P., Wong, K., McQuaid, J., Farrell, B., Preston, L.A., Tan, X., Keller, M., Mathur, E., Kretz, P.L., Burk, M.J., and Short, J.M.			
TITLE	Direct Submission			
JOURNAL	Submitted (26-NOV-2003) Bioinformatics, Diversa Corporation, 4955			
REFERENCE	Submitted (26-NOV-2003) Bioinformatics, Diversa Corporation, 4955			
AUTHORS	Directors			
TITLE	Location/Qualifiers			
JOURNAL	1..1014			
FEATURES	/organism="uncultured organism"			
source	/mol_type="genomic DNA"			
	/db_xref="taxon:155900"			
	/clone="2A7"			
	/environmental_sample			
	1..1014			
	/locus_tag="BD5275"			
	1..1014			
	/locus_tag="BD5275"			
	/EC_number="3.5.5.7"			
	/codon_start=1			
	/evidence=experimental			
	/transl_except="(pos:1..3,aa:Met)"			
	/product="nitrilase"			
	/protein_id="AAR97396.1"			
	/db_xref="GI:40890103"			
	/translation="MKEAIKVCQVAPFVFLDIAITDKYVALIIEAARNCARLIAPP"			

ETTITPCWPFMTLDDPAMQMGQFVRVYHNSLVLDSPOAKRISSEAQAQAGIYVALGYSB  
 RVSGTITPMFGOMLDDDKCETAGLRKRLKTPHVBERTLPEEGGSSLSTEDTPIGLVJGJLC  
 CMWELDITSKALYALVHETFPFAMPSPFISYRQATVEVTLGEVWVAARISYVAEBOCCAVT  
 LASALYSPENITMILCNDSKSKSLDAGGGSYRSLIGDGSVLNAPLCENEGIGIYATITL  
 DPFARITYAKTADDPAGHYRSPDVTRLILNSANQPVVEGRELDPASAQGEVEEAAPGV  
 GGD<sup>10</sup>

Query Match	49.1%;	Score 497.8;	DB 3;	Length 1014;
Best Local Similarity	68.2%;	Pred. No. 1.1e-136;		
Matches 691;	Conservative 0;	Mismatches 322;	Indels 0;	Gaps 0;

[illegible]



```

CDS
    locus_tag="Peyr_0002"
    1575..2678
    /locus_tag="Peyr_0002"
    /EC_number="2.7.7.7"
    /codon_start=1
    /evidence=not experimental
    /transl_table=11
    /product="DNA polymerase III, beta chain"
    /protein_id="AA135076.1"
    /db_xref="InterPro:IPR01001"
    /db_xref="GI:63253980"
    /translacion="MHFTIOREALKLPOLVAGVERROTLPVLSNVLLVVEGQQLSL
    TGDLEIVGVRVQLERPARGETTPARKMDICTSINDALIDIDLDSKILVKKG
    RSRFTLTPANDPPTVEBPGSLTFVLVQSKRLIRISFMAQDVRYLYNGMLL
    EVSAGILRAVATDGHRLKMSADITHADHQVTPRKGLLEAKLLTEDGVSTV
    LQGHRIATGTEFTFTSKLVGKPPDYERVLPGKGDVLGDRQALREAFSRALISN
    EKYGRILQLASGQLKQANNPEOEBAEEISVEYNDLSLEIGNVSYYLLDVLGVMTT
    EOVRLISDSNSALVQESDNDSDAAYVMPRL"
    2701..3804
    /locus_tag="Peyr_0003"
    2701..3804
    /locus_tag="Peyr_0003"
    /codon_start=1
    /evidence=not experimental
    /transl_table=11
    /product="RecF protein"
    /protein_id="AA135077.1"
    /db_xref="GI:63253981"
    /db_xref="InterPro:IPR01238"
    /db_xref="InterPro:IPR003395"
    /translacion="MSLSRVSTGVNMLPVTLSPSPRINILYGANGSKTSLVLAH
    LIGTARSRSRLPVIOYRQPSCTVQVNDLQAGHSNLGVSRRDGEPIRIDGON
    ARSAAQALRILPIQLINPDSFRLLBGA PKIRQGLDGVFVVERPAAWRLOKALK
    QRSNMLHGTLDAAQAMRELSASDEIFRAIYKALKVEQTLSELVLEGL
    TLSYRGMDKEKELSTVLASSIHRDQGMHQAGPORAADRLGAVHNAADLSRGQO
    KLVCALRIAGHLVSQVRGQCTYLVDLPSRLDMDHRRALCRLLBELRCQVFTCV
    DOEFLREGMOTETPEALFHVHGRIYTHDHRH"
    3809..6226
    /locus_tag="Peyr_0004"
    3809..6226
    /locus_tag="Peyr_0004"
    /EC_number="5.99.1.3"
    /codon_start=1
    /evidence=not experimental
    /transl_table=11
    /product="DNA gyrase, B subunit"
    /protein_id="AA135078.1"
    /db_xref="GI:63253982"
    /db_xref="InterPro:
    IPR000565:IPR001241:IPR002288:IPR003594:IPR006171:IPR01155
    7:IPR011558"
    /translacion="MSNQYDSSSIKYLKGLDAVRKRGMYIGDGDGSGLHMVFE
    VVNSIDBALAGHCDDDISIIHPBESTVVRNNGIIPDVNKEBVSAAEYIVLVA
    GGRPDNSYKVSGLHGVSVNALSHLLLTVRSGKIWEQTVHGVPOEPKVLIG
    ESDGTQIOHFKPSAEIPKNIHFSMDILAKIRIELSPLNSGVGLXKDERGKELPQ
    YEGGLRAVEYLNNTKTPVNEVPHFNIDRDGIGVEILQWNSFNEMLCFTNNIPQ
    RDGTHLVGFRSALTNNLNNYIEOGGLAKKVAATGDDAREGLTALISVAPPKPS
    SOTRDXLVSSEVTAVRQEMKYSFDELLENPAKAVGMDIAAARBAARAKRM
    TRKRGALDIAGLPGLADCCQKDPALSTLYVEGDSAGSAGKQGNRTQAILPLKKG
    ILNVEKAFDPMISSQEVGTLITAGCGIGREEVNIDLRHNHIIIMTDAVDSSHIR
    TLLLPFRQLPELIERGYIYIAQPLPKVKGQEOYIKXDEAMEBYMTQSALEDAS
    LHSESAAGISGTALRKLNDPRMVKTLKLSRLYPOELTEHFVEYLPITLEQSLDH
    EGMQAMLAIPARLRGTGKGLVYKASIREIRERNVMLPEVELISHGLSNVTFNRDP
    FGSNDKYTVTLAGAQLSTLEBGAIVQGEKKKPVNEKELANLMAESTRHTIIGRY

gene
CDS
    locus_tag="Peyr_0004"
    3809..6226
    /locus_tag="Peyr_0004"
    /EC_number="5.99.1.3"
    /codon_start=1
    /evidence=not experimental
    /transl_table=11
    /product="DNA gyrase, B subunit"
    /protein_id="AA135078.1"
    /db_xref="GI:63253982"
    /db_xref="InterPro:
    IPR000565:IPR001241:IPR002288:IPR003594:IPR006171:IPR01155
    7:IPR011558"

Query Match      42.1%  Score 427.4  DB 1  Length 110000;
Best Local Similarity 65.9%  Pred. No. 1.3e-115;
Matches 620; Conservative 0; Mismatches 321; Indels 0; Gaps 0;

QY 1 ATGAAAGAGCTATCAAGTGCCTGCGTGAAGCCGCCGATCTACATGATTTGGAG 60
DB 9243 ATGAAAGAGCTTGTGAAGTGCCTGCGTGAAGCCGCCGATCTCTGATCTGAC 9184
```

```

QY 61 GCCACGGTGGACAAACCATTTGATGTGAAGAGACACAGTAAATGCTCGTCTG 120
DB 9183 GCACCCGTAGACAAAGACATTAACCTTGATGAGCAGCGGACGCCGAGCCGATCTT 9124
QY 121 ATGCGCTTTCCGAAACCTTGATTCAGGCTACCATGATGTTTCTTTGGCTGATCA 180
DB 9123 ATTGCATTCGCCGAAACGTGGATTCGGGGCTATCCGGTGTTTGTGTGCTGACGAC 9064
QY 181 GCATGGCAATGCATATTTGTAACGCCCAATACCATGAGAACATCATTTGAGTGG 240
DB 9063 GCGTGGAAACATGCGCTTGTCAGCGCTTATCACACAGATCTCTGTCGTGGACA 9004
QY 241 CAAGCTAAGCGCAATTCAGATGACGCAACGCGTGGGAATCATGTGACCTTGGGAT 300
DB 9003 CAGGCCGAGATGATCAGCATGCGGCTCGGCATCTCGTCTCTATGTGTGCTGGTTAC 8944
QY 301 AGTGAACGGGTGCGTGGACACCTTTACATGATGATGATGTTCATGAGCATATG 360
DB 8943 AGCAGCGCAACAAAGCGTGCCTGTATATCGTCAAGTATCATCGATCACGGTGAA 8884
QY 361 ACCATTTGGGCGCGCGCAAGTTGAACCTAATTTGTTGAAGTACTTTTGGGGGAA 420
DB 8883 ACGGTGGCGTGTGCTGCAAGCTTCAAGCTATCATGTGAACGTACGATGTTGGGAA 8824
QY 421 GGGGATGTTCAATGCTAGCGGTTTTCGAGACGTCTGTGGAAAGCTGGTGCTTATGC 480
DB 8823 GCGGACGGGCAATCATTTGCGTACCTTTGAAACCCCTGTGCGGCTTCTCGGCGCTT 8764
QY 481 TGTGGAGACACTTCAACCGCTTAACAAATACGCTTTGTATGACAAATGAAGATT 540
DB 8763 TGCTGGAGAGACCTGCAACCGCTTTCCAAGTACCGCATGTACGACGACGAACAT 8704
QY 541 CATGTGCGGCTGGCCGAGCTTTAGCCTTATCTCATGCGGGAAGCCCTGGGGCT 600
DB 8703 CACGTACCGGCTGGCCCAAGCTTACGCTGACCGCAACGCAAGATGCTTGGGCC 8644
QY 601 GATGTCAATGATGAGCGGCTCTCGAATCTATGCCGTGAAGGCAATCTTGTAACG 660
DB 8643 GAGGTCAATACGTGGGCTTGGCGGCTTATGCGCAGAAAGGCAATGCTTGTCTGCA 8584
QY 661 TCGTGTGCGCTGTTTCAATCATATGATGATGATGCTTTGTACAGATGACGAAAG 720
DB 8583 CCTGTGCAATCGTTTCCCTGAAATGATGAAATGCTTTGCGACAGTGAAGCCAA 8524
QY 721 GCGTGTGCTGCGTGGATGATGACACTCACTGATCATTAGAGGCTGATGATGAT 780
DB 8523 AGCCTGTGACGAGCGCGGTGGCGGACGACGAGATTTTCGTCGGACGAAAGTAT 8464
QY 781 GTCGCGGCTTGTGCGCAAAATGAAGAGGATATCTTACGCAACCTTGATCTGGA 840
DB 8463 GCGACCCCATTTGGCGCAACATGAGGAAGTCTGCTTACGCAACCTTGATCTGAG 8404
QY 841 CGCATCTTGTAAAGTGGCGGACCTGCTGCTATTAATCCGTCGGCAATTAAT 900
DB 8403 TTGACCTTGGCAAGTAGACGCGGACCTGCTGACATTAATCAACGTCGGCACT 8344
QY 901 CGCTTGCTAATAGATCGACGCGCTTAATTAATCCGCTAGTTGA 941
DB 8343 CGTCTGATTTCAATCCCAATCCCACTCCATGTTGATGCA 8303

RESULT 6
AE016853_02
WPCOMMENT
Sequence split into 64 fragments LOCUS AE016853 Accession AE016853
Fragment Name Begin End
AE016853_00 1 110000
AE016853_01 100001 210000
AE016853_02 200001 310000
AE016853_03 300001 410000
AE016853_04 400001 510000
AE016853_05 500001 610000
AE016853_06 600001 710000
```



AE016853\_07 700001 810000  
AE016853\_08 800001 910000  
AE016853\_09 900001 1010000  
AE016853\_10 1000001 1110000  
AE016853\_11 1100001 1210000  
AE016853\_12 1200001 1310000  
AE016853\_13 1300001 1410000  
AE016853\_14 1400001 1510000  
AE016853\_15 1500001 1610000  
AE016853\_16 1600001 1710000  
AE016853\_17 1700001 1810000  
AE016853\_18 1800001 1910000  
AE016853\_19 1900001 2010000  
AE016853\_20 2000001 2110000  
AE016853\_21 2100001 2210000  
AE016853\_22 2200001 2310000  
AE016853\_23 2300001 2410000  
AE016853\_24 2400001 2510000  
AE016853\_25 2500001 2610000  
AE016853\_26 2600001 2710000  
AE016853\_27 2700001 2810000  
AE016853\_28 2800001 2910000  
AE016853\_29 2900001 3010000  
AE016853\_30 3000001 3110000  
AE016853\_31 3100001 3210000  
AE016853\_32 3200001 3310000  
AE016853\_33 3300001 3410000  
AE016853\_34 3400001 3510000  
AE016853\_35 3500001 3610000  
AE016853\_36 3600001 3710000  
AE016853\_37 3700001 3810000  
AE016853\_38 3800001 3910000  
AE016853\_39 3900001 4010000  
AE016853\_40 4000001 4110000  
AE016853\_41 4100001 4210000  
AE016853\_42 4200001 4310000  
AE016853\_43 4300001 4410000  
AE016853\_44 4400001 4510000  
AE016853\_45 4500001 4610000  
AE016853\_46 4600001 4710000  
AE016853\_47 4700001 4810000  
AE016853\_48 4800001 4910000  
AE016853\_49 4900001 5010000  
AE016853\_50 5000001 5110000  
AE016853\_51 5100001 5210000  
AE016853\_52 5200001 5310000  
AE016853\_53 5300001 5410000  
AE016853\_54 5400001 5510000  
AE016853\_55 5500001 5610000  
AE016853\_56 5600001 5710000  
AE016853\_57 5700001 5810000  
AE016853\_58 5800001 5910000  
AE016853\_59 5900001 6010000  
AE016853\_60 6000001 6110000  
AE016853\_61 6100001 6210000  
AE016853\_62 6200001 6310000  
AE016853\_63 6300001 637126

Continuation (3 of 64) of AE016853 from base 200001 (AE016853 Pseudomonas syringae pv. b)

Query Match 40.7% Score 412.8; DB 1; Length 110000;  
Best Local Similarity 64.8%; Pseud. No. 2.9e-111;  
Matches 612; Conservative 0; Mismatches 332; Indels 0; Gaps 0;

QY 1 ATGAAAGAGCTATCAAGTCCTGCGTGAAGCGCCCGCATCTACATGATTTGGAG 60  
DB 12501 ATGAAAGCGCCACTGAAGTTCCTGCGTGAAGCGCCCGCATCTGATCTGAGC 12560  
QY 61 GCGACGCTGACAAACCATTTGATGAGAAAGACGACGCTAATATCTCTGCTG 120  
DB 12561 GCGACGCTGACAAACCATTTGATGAGAAAGACGACGCTAATATCTCTGCTG 12620  
QY 121 ATGCGCTTCCGGAAGTTGATTCAGGCTACCATGGTTCTTTGGCTTGACTACCA 180

DB 12621 ATTCATTCCTCCGAGACGTGATTCGCGGCTACCACTGTTTTTGTGTGCTGACGACCG 12680  
QY 181 GCATGGCAATGCAATTTGTATGCGCAATACCATGAGAACTCATTTGAGTGAATGGCCCT 240  
DB 12681 GCTGGAACATGCTCCGTTGTGATCATGCTATACCAACAGTCTCTGTGTCTGACAGTCT 12740  
QY 241 CAAGCTAAGCGCATTTTCAGATGACAGCCAGCGGTTGGAAATCATGTCTACCCCTGGGATG 300  
DB 12741 CAGGACGAGCATCATGATGACAGTCTGTGACGACGCTATCTACGTGTCTGGTTAC 12800  
QY 301 AGTGAACGAGTGTGGTGGGACCTTTTACATGATGATGTTTCATAGGCGAATATGTATAC 360  
DB 12801 AGCAGCGCACGAAAGGCTTCCCTGTATATCGTACAGTATCATGATCAAGCGGACG 12860  
QY 361 ACATTTGGGGCGCGCGCAAGTTGAAACCTTCTTTTGTGAACGATCTTTGGCGGCA 420  
DB 12861 ACCGTGGACGCGCGCGCAACCTCAAGGCGACGATGGAACGTACATTTTGGCGAA 12920  
QY 421 GGGATGTTTCATGCTAGCGGTTTTCAGACGTCTGTGGAAGGCTGGGTATGTC 480  
DB 12921 GGAACGCGGCGGTATTTGCGACCTTTGAAACACTGTAGGCGTACTGGTCCCTTTCG 12980  
QY 481 TGTGGAGACCTTTCACCGCTTAACAAATACGCTTTGTATGCAAAATGAAGATTT 540  
DB 12981 TGCTGGAGACCTTTCACCGCTTTCGAAATACGCTATGACGCGCAGAAACGATTT 13040  
QY 541 CATGTGCGCTTGGCGGAGCTTATGCTTATCTATGCGGGAAGCCCTGGGCGCT 600  
DB 13041 CACGTAGCGCTGCGCCAGCTTACGCTTGAACGACGACCGCGCGCTTGGCCCA 13100  
QY 601 GATGTCATGTAAGGCGGCTTCGATCTATGCTGTTGAAGGCGCAATGCTTGTACTAGCG 660  
DB 13101 GAGGTCAACACCGCGCTTCACGCTTATGCGCAGAAAGGCGAGTGTGTCTATGCA 13160  
QY 661 TCGTGTGCGCTGTTTCAATCATGATGATGATGCTTTGTAGACATGACAAAGCAT 720  
DB 13161 CCGTGCAGATGATCTCTCCGGAATGATGAGATGCTGTGACAGGATGCCAAGCAC 13220  
QY 721 GCGTTGCTTGTGTGTGTGTGTGACATCATAGTATCATAGGCGCTGATGTGTGACTTG 780  
DB 13221 AGCTTGTGACGCGCGGCGGAGGATGACGAAATCTTTGGCCGCGAGGAGGACCTTC 13280  
QY 781 GTGCGGCTCTTGGCGAAATGAGAGGGATTTCTCTACGCAAACTTGATCTGAGATA 840  
DB 13281 GCGACCCGCTTGGCGGAGATGAGAAAGGCTGCTTACGCAACTTGATCTGCGCT 13340  
QY 841 CGCATCTTGTAAATGCGCGACAGCCCTGCTGATTAATTCCTGCTCCGACATTAAT 900  
DB 13341 TTGATCTTCCGCAAGTGTGCGGCGGACCGCGCAGGAGCATCTGCGGCTGAGCTACA 13400  
QY 901 CGTTTGTATATGATGCGAGCCCTTAATTAATTAATTAATTAATTAAT 944  
DB 13401 CGCCTGATGTTCAACCCCAATCCCAACCTTTCGCTGTGTGAAT 13444

RESULT 7  
AY487512 999 bp DNA linear ENV 05-APR-2004  
LOCUS AY487512  
DEFINITION Uncultured organism clone 2A9 nitrilase (BD5168) gene, complete  
cde.  
ACCESSION AY487512  
VERSION AY487512.1 GI:40890228  
KEYWORDS ENV.  
SOURCE uncultured organism  
ORGANISM unclassified; environmental samples.  
REFERENCE 1 (bases 1 to 999)  
AUTHORS Robertson,D.E., Chaplin,J.A., Desantis,G., Podar,M., Madden,M.,  
Ch,E., Richardson,T., Milan,A., Miller,M., Weiner,D.P., Wong,K.,  
McQuaid,M., Farwell,B., Preston,L.A., Tan,X., Shedd,M.A.,  
Keller,M., Mathur,E., Kretz,P.L., Burk,M.J., and Short,J.M.  
TITLE Exploring nitrilase sequence space for enantioselective catalysts  
JOURNAL Appl. Environ. Microbiol. 70 (4), 2429-2436 (2004)

REFERENCE 2 (bases 1 to 999)  
AUTHORS Robertson,D.E., Chaplin,J.A., Desantis,G., Podar,M., Madden,M.,  
Ch,E., Richardson,T.H., Milan,A., Miller,M., Weiner,D.P., Wong,K.,  
McQuaid,J., Farwell,B., Preston,L.A., Tan,X., Keller,M., Mathur,E.,  
Kretz,P.L., Burk,M.J. and Short,J.M.  
TITLE Direct Submission  
JOURNAL Submitted (26-NOV-2003) Bioinformatics, Diveresa Corporation, 4955  
Director's Place, San Diego, CA 92121, USA  
FEATURES  
source  
1..999  
/organism="uncultured organism"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:155900"  
/clone="2A9"  
/environmental\_sample  
1..999  
/locus\_tag="BD5168"  
1..999  
/locus\_tag="BD5168"  
/EC\_number="3.5.5.7"  
/codon\_start=1  
/evidence=experimental  
/product="nitritease"  
/protein\_id="AAR97459.1"  
/db\_xref="GI:40890229"  
/translation="MKTTVTVAACVQAAPVFMDEGTIDKTTTILISEAOKGAEILAF  
ETWIPGYWFLMNSPATNMLVVOYHNSLVLDSTOKRIADAAKONNITVALGFSE  
RDHSIYIAOWLIGSDGETIGIRRKIKATHTERTLFGSDSSITWTMTPTGANGALC  
CWEHIOPLSRMYMSQHEHIIHAMPSSITSYTAALGPVNTASLTYAABQCPY  
IAPCAVSDENIDILCPDDRRALISAGGAKIYGPGRGLVPLGENEGELIAEL  
D5AAITPAKLADPVGHYSRDPYTRLLFPNSANKTVIKRHSPEELIAEQIAEEEEE"  
ORIGIN  
Query Match 39.4%; Score 399.6; DB 3; Length 999;  
Best Local Similarity 64.0%; Pred. No. 2,1e-107;  
Matches 603; Conservative 0; Mismatches 339; Indels 0; Gaps 0;  
QY 1 ATGAAAGAAGTATCAGTGCCTGCTGTCAGACGCCGCCGATCTACATGATTTGGAG 60  
DB 1 ATGAAACAACGGTTACCGTTGCCCTGCGTTCAAGCGCCCGATTTATGATTTAGAA 60  
QY 61 GCGACGGTGAACAAACCATTTGATGTGATGAGAAGACAGACCGTAATATGCTGCTG 120  
DB 61 GCGACCGTATGATTAACATCACTCTCTGTAAGCGGACAGAAAGCGCGAGCTC 120  
QY 121 ATCGCTTTCCGGAACCTTGATTCAGAGCTAACCCATGATTTCTTTGGCTTGACTACA 180  
DB 121 ATCGCTTTCCGGAACCTTGATTCAGAGCTAACCCATGATTTCTTTGGCTTGACTACA 180  
QY 181 GCATGGGCAATGCAATTTGTACGCCAATACATGAAACTATTGAGTTGGATGGCCCT 240  
DB 181 GCCCAAAATATGCCCCCTGTTATATGATATATCAAGAACTCTGTGCTCGACAGTACC 240  
QY 241 CAAGCTAAGCGATTTGATGATGACGACGAGCGGTTGGGAATGATGTCACCTGGGGAG 300  
DB 241 CAGGCGAAGCGAATTTGCGGATGCGGACAGGAATATACATCACTGTGCTTGGGCTTC 300  
QY 301 AGTGAACGGTGTGGTGCACCTTTTACATCAATCATGATGATGATGCGCATATAGTGAC 360  
DB 301 AGCGAACCGATTCATGAGAGGCTCTATATGCGACAGTGGCTGATGGACGACGCGGAG 360  
QY 361 ACCATTTGGGGCGCGGCGAAAGTTGAAACCTATTGTTGAAAGTATTTGTTGGCGAA 420  
DB 361 ACCATTTGGCATTCGGCGCAAGCTCAAGGCCAGCAGCTGAGACGTAACGCTGTTGGCGAA 420  
QY 421 GGGGATGGTTATATCGCTAGCGGTTTTCGAGACGTCGTGTGGAAGGCTGGTGGCTTATGC 480  
DB 421 AGCGACGGCTCTCCCTGACCACTGGGAGACACTCTGGGTAAGCTCGGGGCGCTTGC 480  
QY 481 TGTGGGAGCACTTCAACCGCTAACAAATATGCTTTGTATGCAAAATGAAAGATT 540  
DB 481 TGCTGGGAGCACTGCAACCGCTGTCCGCTATGCAATGTATTTCCACACATGAGGATC 540

QY 541 CATTGTCGGCTTGGCCGACCTTACGCTTATCTTAATGCGGAAAGCCCTGGGCT 600  
DB 541 CACATCTCTCGCCCGCAGTTTACGCTTACACCGATCAGACGCGGACCTGGGCT 600  
QY 601 GATGTCAATATGAGCGGCTCTGCAATCTATGCGCTTGAAGGCGCAATCTTGATAGCG 660  
DB 601 GACGTCAATATGAGCGGCTCTGCAATCTATGCGCGCGGAGGCGAGTCTTGATAGCC 660  
QY 661 TCGTGTGGCTCGTTTACAAATCATATGATGATGCTTTGTACAGATGACGAAAGCAT 720  
DB 661 CCGGTGCGCGTGTCTTGTGATGAATGATGATTTACTCTGTCTGTGATGACCGAGA 720  
QY 721 GCGTGTCTTGTGCTGTGTGAGACATCACTAGTATCAATAGGCGCTGATGCTTG 780  
DB 721 GCGTGTCTTGTGCTGTGTGAGACATCACTAGTATCAATAGGCGCGGAGGAGATC 780  
QY 781 GTGCGGCTCTTGTGCGGAAATGAAAGGATTTCTTACGCAACCTTGATCTGAGTA 840  
DB 781 GTCAACCCCTCTCGGGGAAATGAGAGGACCTGTATCTGCTGAGCTGAGCTGCG 840  
QY 841 CGCATCTTGTATAAATGAGCGGACGACCTGTGTATATTCGCTCCGACATTA 900  
DB 841 ATTACCTTGTCAAACTGGCGGACGACCGGTTGGCACTATTCGCTGAGCTGAGCTGACC 900  
QY 901 CGCTGTAAATGATTCGACGCTTAATTAACCGGTAGTTGA 942  
DB 901 CGCTGTAAATGATTCGACGCTTAATTAACCGGTAGTTGA 942  
RESULT 8  
AY487492  
LOCUS 1002 bp DNA linear ENV 05-APR-2004  
DEFINITION Uncultured organism clone 2A10 nitritease (BD7615) gene, complete  
cde.  
AY487492  
VERSION AY487492.1 GI:40890188  
KEYWORDS ENV.  
ORGANISM uncultured organism  
SOURCE uncultured organism  
REFERENCE 1 (bases 1 to 1002)  
AUTHORS Robertson,D.E., Chaplin,J.A., Desantis,G., Podar,M., Madden,M.,  
Ch,E., Richardson,T.H., Milan,A., Miller,M., Weiner,D.P., Wong,K.,  
McQuaid,J., Farwell,B., Preston,L.A., Tan,X., Sneed,M.A.,  
Keller,M., Mathur,E., Kretz,P.L., Burk,M.J. and Short,J.M.  
TITLE Exploring nitritease sequence space for enantioselective catalysis  
JOURNAL Appl. Environ. Microbiol. 70 (4), 2429-2436 (2004)  
REFERENCE 2 (bases 1 to 1002)  
AUTHORS Robertson,D.E., Chaplin,J.A., Desantis,G., Podar,M., Madden,M.,  
Ch,E., Richardson,T.H., Milan,A., Miller,M., Weiner,D.P., Wong,K.,  
McQuaid,J., Farwell,B., Preston,L.A., Tan,X., Keller,M., Mathur,E.,  
Kretz,P.L., Burk,M.J. and Short,J.M.  
TITLE Direct Submission  
JOURNAL Submitted (26-NOV-2003) Bioinformatics, Diveresa Corporation, 4955  
Director's Place, San Diego, CA 92121, USA  
FEATURES  
source  
1..1002  
/organism="uncultured organism"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:155900"  
/clone="2A10"  
/environmental\_sample  
1..1002  
/locus\_tag="BD7615"  
1..1002  
/locus\_tag="BD7615"  
/EC\_number="3.5.5.7"  
/codon\_start=1  
/evidence=experimental  
/product="nitritease"  
/protein\_id="AAR97439.1"  
/db\_xref="GI:40890189"  
/translation="MKTTVTVAACVQAAPVFMDEGTIDKTTTILISEAOKGAEILAF

PETWIPGYWPLMNSPATMPLVYQYHONSILVDSQAQKRIADAQONNITVVLGFS  
ERDHSLSYISOWLIGSDSETTIRKRLKATHERITLFGSDSSLTJMETPLGANGAL  
CMEHLIOPISRYAMSQHEIHIAMPSISLYTSALGPDVNTAASRIYAABGOC  
VLAFCAYVSDMDIDLCPDDDRALLSAGGKARIYGPDRGLVPLBENEGGLTIAE  
LDSAATITPAKLADPVGHSRDPVTRLFNPSANKVIKRSHPPLIAEQAPEEEEE"

Query Match 39.1%; Score 396.4; DB 3; Length 1002;  
Best Local Similarity 63.8%; Pred. No. 1.9e-106;  
Matches 601; Conservative 0; Mismatches 341; Indels 0; Gaps 0;

1 ATGAAAGAGCTATCAAGTCCCTGCTGCAAGCCGCCGATCATGATTTGGAG 60  
4 ATGAAAGAGCTATCAAGTCCCTGCTGCAAGCCGCCGATCATGATTTAGAA 63  
61 GCGACGGTGGACAAACCATTAATGATGAGAAAGACGACGTAATATCTGCTG 120  
64 GGCACCATATGATTAACGATCAACCTCATCTCTGAAGCCGACAGAAAGCCGAGCTC 123  
121 ATGCGCTTCCGGAAACTTGGATTCAGGCTACCATGCTTCTTTGGCTTACACCA 180  
124 ATGCGCTTTCGGAGACTGGATACCCGGTACCCGTTCTTATGCTGAACCTGCCC 183  
181 GCATGGGCAATSCAATTTGTACGCCAATACATGAACTCATTTGATGATGAGCCCT 240  
184 GCGAACAATATGCCCTCGTTTATCATATCATCAAGACTCTGTGTCTGACAGTGC 243  
241 CAAGCTTAAGCGCATTTCAAGTACAGCCGATGCGGTTGGGATCATGATGATG 300  
244 CAGCGGAAGCGAATTCGGATGCTGACAGCAATTAACATCATCTGCTTGGGATTC 303  
301 AGTGAACGGGTGGGAGCACTTCAATCATCATGCTGCTTATGAGGATTAATGTAC 360  
304 AGCGAGCGGATCATGAGAGCTCTATATCTCAAGTGTGATGAGCGACGCGAGAA 363  
361 ACCATTTGGGGCCCGGCGAAGTTGAACCTATTTTGTGAACGTAATTTTGGCGCA 420  
364 ACTATTTGGGATTCGGGCGCAAGCTCAAGGCGACACAGTGGAGCGTATGTTGGCGAA 423  
421 GGGGATGGTTCATCGCTAGCGGTTTTCAGACGCTGTGTGGAAAGCGTGGGTATGC 480  
424 AGCGACGGCTCTCCCTGACCACTGGGAGACACCTCTGGGATTAATGCGGCGCTGCG 483  
481 TGTGGGAGCACTTCAACCGCTTACAAATACGCTTGTATGACAAATTAAGAGATT 540  
484 TGTGGGAGCACTTCAACCGCTTGTCTGCTATGAGATGATTTCCAGCATGAAGATC 543  
541 CATTTGGCGGCTTGGCCGAGCTTTAGCTTTATCTTAATGCGGAGAAAGCCCTGGGGCT 600  
544 CATATCGCTGCGTGGCCGAGCTTCACTTATACACAGGCAACTGCGCGCTCGGACCT 603  
601 GATGTCAATGTAAGCGGCTCTTCAATCTATGCGGTTGAAGGGCAATGCTTCTACTAGC 660  
604 GACGTCAACAGCGGCGCTTCAAGCTTATGCGGAGGGGCACTGCTTGTGTAAAC 663  
661 TCGTGTGGGCTGCTTCAACATCATGATCATGATGCTTTGTACAGATGACGAAAGCAT 720  
664 CCATGTGCGGTGGTTTCTGATGAGATGATGATTTACTCTGTCTGATGATGACCGGAGA 723  
721 GCGTGTGCTTGGGCTGATGATGACATCAGTATCATAGAGGCGGATGATGATGATG 780  
724 GCGTATCTCAAGTGGCGGAGGAGACATCCGATTTTACGACCTGATGAGAAAGACATC 783  
781 GTCGCGCTCTTGGCGGAAATGAAGAGGATTTCTACGCAAACTTGATCTTGAGATA 840  
784 GTACACCTCTCTCGGGGAAATGAGAGGACATGCTTATCGCTGAGCTCTGCTGCG 843  
841 GCGATCTCTTGAATAGCGGAGACATCTGCTGATCATTTCCGCTCCGACATTAAT 900  
844 ATCACTTTTGGCAAACTGCGGCAAGATCCCGTAGGCACATTTCCCGCTGACGTGAC 903  
901 CGCTGTAAATGATGACGAGCCCTAAATTAACCGGATGTTGAA 942

Db 904 GCGCTCTTTTAAATCCTTACGCCAACAAGACTGATTAATA 945

RESULT 9 1038 bp DNA linear ENV 05-APR-2004  
LOCUS AY487427  
DEFINITION Uncultured organism clone 2A12 nitrilase (BD7611) gene, complete  
cde.

ACCESSION AY487427  
VERSION AY487427.1 GI:40890058  
KEYWORDS ENV.

SOURCE uncultured organism  
ORGANISM uncultured organism  
unclassified; environmental samples.

REFERENCE 1 (bases 1 to 1038)  
Robertson,D.E., Chaplin,J.A., Desantis,G., Podar,M., Madden,M.,  
Chi,E., Richardson,T., Milan,A., Miller,M., Weiner,D.P., Wong,K.,  
McQuaid,J., Farwell,B., Preston,L.A., Tan,X., Sneed,M.A.,  
Keller,M., Mathur,E., Kretz,P.L., Burk,M.J. and Short,J.M.

Exploring nitrilase sequence space for enantioselective catalysis  
Appl. Environ. Microbiol. 70 (4), 2429-2436 (2004)  
2 (bases 1 to 1038)  
Robertson,D.E., Chaplin,J.A., Desantis,G., Podar,M., Madden,M.,  
Chi,E., Richardson,T.H., Milan,A., Miller,M., Weiner,D.P., Wong,K.,  
McQuaid,J., Farwell,B., Preston,L.A., Tan,X., Keller,M., Mathur,E.,  
Kretz,P.L., Burk,M.J. and Short,J.M.

Direct Submission  
Submitted (26-NOV-2003) Bioinformatics, Diversea Corporation, 4955  
Directors Place, San Diego, CA 92121, USA

FEATURES  
source location/Qualifiers

1..1038  
/organism="uncultured organism"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:155900"  
/clone="2A12"  
/environmental\_sample  
1..1038  
/locus\_tag="BD7611"  
1..1038  
/locus\_tag="BD7611"  
/BC\_number="3.5.5.7"  
/codon\_start=1  
/evidence=experimental  
/product="nitrilase"  
/protein\_id="AA097374.1"  
/db\_xref="GI:40890059"

/translation="MAIEHPRYVAOVAPEFLNLEATVDLTALIEBARGASLI  
AFPTWIPGYWPLMNSPATMPLVYQYHONSILVDSQAQKRIADAQONNITVVLG  
FSEPDAGSLYIAQAILSPBGTITATRKLRKTHYERARFGSDSDLAHVHTKGRVG  
ALCCWEHLQPLSKYAMTAQONQVHIAIMPSTLYVDAYALGPEVNNASRLVAVGQ  
CFVVAFCATVSOQKIMDLCEPQQOALLKPGGHAQIYGPDRSLADLPDADELLY  
ADIDLAATLAKAADPAHGHSRDPVTRLFNPSANKVIKRSHPPLIAEQAPEEEEE"  
EHTLEBGEQA"

ORIGIN

Query Match 35.9%; Score 364.2; DB 3; Length 1038;  
Best Local Similarity 62.2%; Pred. No. 7.7e-97;  
Matches 573; Conservative 0; Mismatches 348; Indels 0; Gaps 0;

18 GGTGGCCGCTGGAGAGCCGCCGATCTAATGATTTTGGAGCGACGGTGAACAAC 77  
27 GGTGGCCGCTTCAAGGGGCGCAAGTTCTTAATCTGAGGCAACCGTTGACAAAG 86  
78 CATTAGTGTATGAGAAGAGCAGTAATATGCTGCTGATGCGCTTTCGGAAAC 137  
87 GATTGCGCTTATGAGGAGGCGCGCCGGAGGCGCAAGTCTCATTTGCTTCCCTGAAC 146  
138 TTGATTTCCAGGCTAACCATGTTTCTTTGGCTTGACTCACACAGATGGGCAATGCAATT 197  
147 TTGATATCTGGAATATCATGTTGCTGCTGCTGGCGCGCATTCGGGGCATGAATTT 206  
198 TGTACGCAATACATGAGAACTCATTTGAGTTGATGGCCCTCAAGCTAAGGCAATTC 257

Db	207	CATCCAGGCGTACCATGACAACTCGATGGTGTATCGACGGCGCCGACGTTGAGCGGATATTCG	266
Qy	258	AGATGACAGCAACGCGGTTGGGATATATGGTCACTCTGGGGATGAGTGAACGGGTGCGTGG	317
Db	267	GCAAGCACACATCGCGCTTGCCAACTCACATCGTTGTGTGGCTTTTACCGAAGAAAGACGACGG	326
Qy	318	CACCTTTACATCACTAGTGGTTCATGAGCGCATATAGGTGACACATTTGGGGGCCCGAG	377
Db	327	AAAGCTGTATATGTGGCCAGGCGCATCTCTAGCCCTTAAGGAAAGACATCGCCACGCGTGG	386
Qy	378	AAAGTTGAACCTACTTTTGTGTGAACGTACTTTGTTGCGCGAAGGGGATGGTTCAATCGCT	437
Db	387	CAAGGTGAACCCCACTCATGTGCGAACGCGCGCATCTTCGCGGAAGGCGACGCGACGACCT	446
Qy	438	AGCGTTTTCGAGACGTCTGTTTGAAGGCTGGGTGGCTTATATGCTTTTGGGAGCACTTCA	497
Db	447	GGCAGTTTCAACACACCAAGCTCGACGGGTGGGCGCCTTTGTCTCTGGAGACATCTTCA	506
Qy	498	AACCGTAAACAAATACGCTTTGTGTATGACAAATAGAAAGATTCATTGTGCGGCTTGACC	557
Db	507	GCCACTTTCCAAATACGCGATGTATGTACAGAACAGCAGAGGTCCATATGCGCGGCTGGCC	566
Qy	558	GAGCTTTAGCCTTATTCCTAATACGGCGCAAAAGCCCTGGGGCCTGATGTCAATGTAGCGCG	617
Db	567	CAGTTTTCCTTTACGTGACGCGGCTTACCGGCTTGGGCGAAGGTGAACAACGCGCG	626
Qy	618	CTCTCGAATCTATGCCGTGTGAAGGGCAATGCTTCTGTACTAGCGTGTGTGCGCTGTTTC	677
Db	627	GAGTGGTGTATGGGTCGAGGGCGAAGTGCTTCGTGGTGTGGCGCGTGTGCAACGGTTTC	686
Qy	678	ACAATCCATGATTCGATATGCTTTGTATACAGATGACGAAAAGCATGTGGTTCCTTGGCTGG	737
Db	687	GCAAAAGATGATCGATGTGCTTTGGGAGACACCCGAGCAACAAAGGCTCTTTGMAACCGGAG	746
Qy	738	TGTTGGAACATCACTATCATAGGCGCTGATGTGTGTGACTTGTGTGCGGCTCTTGCGGA	797
Db	747	GGGTGTGACGCGCAAAATCTACGGTCCGACGGAAGATCGTGGCGAATCGGCTGCTGCC	806
Qy	798	AAATGAAGGGTATTTCTCTACGCAAACTTTATCTTGAGATGACGATCTTTGCTAATAAT	857
Db	807	GGACGCGGAGGGGCGTTGTATGACGAACATTGACCTTGAGGCATCACCTTCGCAAAAGC	866
Qy	858	GACGCGACAGCCGTGCTGCTATTTATTCGCGTCCGACATTACTGCTTGTAAATGATGCG	917
Db	867	AGCTGCAATCTGTGTGGCCATTACTCTGCGCCCTGACGTGAACAACACTGCTGTTGACCG	926
Qy	918	CAGCCTTAATTAACCGGTAGT	938
Db	927	CAATCCAAAGCCCGTGTGCT	947

RESULT	10
AY487546	
LOCUS	1011 bp DNA linear ENV 05-APR-2004
DEFINITION	Uncultured organism clone 2A16 nitrilase (BD5295) gene, complete cds.
ACCESSION	AY487546
VERSION	AY487546.1 GI:40890296
KEYWORDS	ENV.
SOURCE	uncultured organism
ORGANISM	uncultured organism
REFERENCE	unclassified; environmental samples.
AUTHORS	1 (bases 1 to 1011)
	Robertson,D.E., Chaplin,J.A., Desantis,G., Podar,M., Madden,M., Ch,E., Richardson,T., Milan,A., Miller,M., Weiner,D.P., Wong,K., McQuaid,J., Farwell,B., Preston,L.A., Tan,X., Sneed,M.A., Keller,M., Mahur,E., Kretz,P.L., Burk,M.J. and Short,J.M. Exploring nitrilase sequence space for enantioselective catalysts Appl. Environ. Microbiol. 70 (4), 2429-2436 (2004) 2 (bases 1 to 1011)
TITLE	
JOURNAL	
REFERENCE	
AUTHORS	Robertson,D.E., Chaplin,J.A., Desantis,G., Podar,M., Madden,M., Ch,E., Richardson,T.H., Milan,A., Miller,M., Weiner,D.P., Wong,K., McQuaid,J., Farwell,B., Preston,L.A., Tan,X., Keller,M., Mahur,E.

**TITLE** Kretz, P. L., Burk, M. J. and Short, J. M.  
**JOURNAL** Direct Submission  
**FEATURES** Submitted (26-NOV-2003) Bioinformatics, Diversa Corporation, 4955  
Directors Place, San Diego, CA 92121, USA  
Location/Qualifiers

```

/organism="uncultured Organism"
/mol_type="genomic DNA"
/db_xref="taxon:155900"
/clone="2A16"
/environmental_sample
1. .1011
gene
1. .1011
locus_tag="BD5295"
1. .1011
CDS

```

```

/BC_number="3.5.5.7"
/codon_start=1
/evidence=experimental
/translate_except={pos:1..3,aa:Met}
/product="nitrlase"
/protein_id="AR97493.1"
/db_xref="GI:40890297"
/translation="MGSHPKFKAIVVQAPFLDDGDGVERAASLIQGAAGAAGAC
APPETWIPGYPMHTWIGSPAMAMEKGFVORFPNALRHGSPQERISGAAEKIKI
LGAERGGCTLYAOWLIGDQOTISRRKIKTHERVYFVGSGDSDSLVDHTAD
TSGLCSEHRLQPLSKYAMVAKONBOQHITGAMPSSLYQPPNALSPENVINAVSVY
GQGFPLPACATVDAMITETCDIPKQGLRGGGHAIIPDGSLLITFVATVITY
LYVELDGVISIAKSAADPAGHYSRDPDITRLLLNQPSSKRVQNMVPLPTEVPEQ
QP"

```

Query Match 31.5%; Score 319; DB 3; Length 1011;  
 Best Local Similarity 60.1%; Pred. No. 2.3e-93;  
 Matches 551; Conservative 0; Mismatches 360; Indels 6; Gaps

QY	1	TCGAAGTTCGCGTGGTGTGCAAGCCGCCCGCCGATCTCAACTGATATTTGGAGGCGCAGCGTGGACA	7
Db	23	TCGAAGCGTGGCGGTGTGTCAGAGCCGGGGCTCGCTTCCCTCGACCTCGACCGGCGGGGTGAAAC	8
QY	74	AAACCATTTGAGTTGATGTGGAAGAAAGCAGCAGCTAATAATGTCGTCTGATGCGCTTTCCGG	13
Db	83	GAGCGCGTGTGCGCTCATGTGGCCAAAGCAGCGCGGCCGAAAGGGGACAGCTGATGTGCTTTCCTG	14
QY	134	AAACTTGGATTTCCAGCGCTACCACTGGTTTCTTTGGCTGTGCTCACCAGCATGGGCAATGC	15
Db	143	AAAGCTGGAATTTCCGGGTACCTCGTGACACCTGGCTTGGCAGCCCGGCGTGGGCAATG	2
QY	194	AA-----TTTGTACGCCCAATACCATAGAACTCATTTGAGTGTGATGCGCTTCAAGCTA	2
Db	203	AAAAAGCTTTGTTCACAGCATATTTGCAACAACCGTGTGGCGGATGGTTCCTCCGCAAGCG	2
QY	248	AGCGCATTTTCAGATGTGACGCGCAAGCGGTTGGGAATCATGTGTACCTTGGGGATGATGTAAC	3
Db	263	AGCCAAATCTCGGGGGCTGGCGCGGAGACAAGATTTATGGTGTGCTTGGGTTTGGCAAC	3
QY	308	GGGTGCGTGGACCGCTTTCATCATGTCAGTGGTTCATAGGCGATTAATGTGACCACTTG	3
Db	323	GCGATGAGGACGCGCTTATATTCGCGAGTGGCTCATCGAACCCGACGCGCAAACTATCT	3
QY	368	GGGCGCGGCGAAGATTGAAACTACTTTTGTGAACTATTTGTTTGGCGGAAGGGGATG	4
Db	383	CAGCAGCGCGGGAAGCTTAAGCCGACTCAGCTCGACGCACTGATTTGGGGGAAGGAGACG	4
QY	428	GTTTATGCGCTAAGCGGTTTTCGAGACGTCCTTGTGGAAGGCTGGGTGGCTTATGCTGTTGGG	4
Db	443	GAAGCGATCTCTCGTGTGATGATACGCGCGCTTGGAGGTATCGGCTCATCTTTGTTGGCTGG	5
QY	488	AGCAGCTTCAACCGCTAACAAAATACGCTTTGTATGCACAAAATGAAGACATTCATGTG	5
Db	503	AGCATTGTGCAACGCTTGTCAAAATACGCAATGTATGCGCCGAATGAAACAGATTTCACATTC	5
QY	548	CGGCTTGGCGAGCTTTAGCGCTTATATCTAATGCGGCGAAAGCGCTTGGGGCTGATGTCA	6

Db	563	GCGCATGGCCAGCCTTTTCGCTATACCAAGCATTTCGCATAAGCGGTAGTCCCAAGTCA	622
Qy	608	ATGTAGCGGACCCTTCGAAITCTATAGCCGTGAAGGGCAAATGCTTCGTACTAGCGTGSTG	667
Db	623	ATATCGCAGTAAGCGGCCGTATACGCCGTGGAAGGCCAATGTTTTCTTCCTCGCGCCGTGCG	682
Qy	668	CGCTCGTTTCACAATCCATGATCCATATGCTTTGTACAGATGACGAAAAGCATGCGTTGC	727
Db	683	CGACGGTTTCGGAACCCCAATGATCGAAAACACTGTGCGATACGCCCGAAAAAGCAGGGACTGA	742
Qy	728	TTCTNGCTGNGTNGTAGACCTCACGATCATPAGGGCCGTATGNTGNTGTACATTGGTCCGCG	787
Db	743	TTCCGCGCGGTATGGCGGGCAGCCGCCGATCTTTCCGCCAGAATGAAGTCTTGCTAGCCCTTA	802
Qy	788	CTCTTGCCGAAATAGAGAAGGGTATTCTCTACGCAAACTTGATCCTGAGATGACATCC	847
Db	803	CGGTAGCGGATATCTTACAGAGGGCCCTCGTATGACGAATCGAACCTCGGGCCTCATTTGCA	862
Qy	848	TTCCTAAAATGGCGGCAACCTCGTGTGTCATTTATCCCCTCCGACATTACTGCGTTGC	907
Db	863	TCCGCAAAAGTGCACGCGACCCCGCGGCACTATTCCGCGGCGAGATGTCAACGCCCTTC	922
Qy	908	TAATGATCGCAGCCCT	924
Db	923	TATTGAATCAGACGCT	939
RESULT 11			
LOCUS	Ay487438	1026 bp	DNA linear ENV 05-APR-2004
DEFINITION	Uncultured organism clone ZA14 nitrilase (BD7266) gene, complete cds.		
ACCESSION	Ay487438		
VERSION	Ay487438.1	GI:40890080	
KEYWORDS	ENV.		
SOURCE	uncultured organism		
ORGANISM	uncultured organism		
REFERENCE	unclassified; environmental samples.		
AUTHORS	1 (bases 1 to 1026)		
TITLE JOURNAL REFERENCE AUTHORS	Robertson,D.E., Chaplin,J.A., Desantis,G., Podar,M., Madden,M., Chi,E., Richardson,T., Milnan,A., Miller,M., Weiner,D.P., Wong,K., McQuaid,J., Farwell,B., Preston,P.L., Burk,M.J. and Short,J.M.		
	Exploring nitrilase sequence space for enantioselective catalysis Appl. Environ. Microbiol. 70 (4), 2429-2436 (2004)		
	2 (bases 1 to 1026)		
	Robertson,D.E., Chaplin,J.A., Desantis,G., Podar,M., Madden,M., Chi,E., Richardson,T.H., Milan,A., Miller,M., Weiner,D.P., Wong,K., McQuaid,J., Farwell,B., Preston,L.A., Tan,X., Keller,M., Mathur,E., Kretz,P.L., Burk,M.J. and Short,J.M.		
	Direct Submission		
	Submitted (26-NOV-2003) Bioinformatics, Diversa Corporation, 4955 Directors Place, San Diego, CA 92121, USA		
	Location/Qualifiers		
	1..1026		
	/organism="uncultured organism"		
	/mol_type="Genomic DNA"		
FEATURES SOURCE	/db_xref="taxon:155900"		
	/clone="ZA14"		
	/environmental_sample		
	1..1026		
	/locus_tag="BD7266"		
	1..1026		
	/locus_tag="BD7266"		
	/EC_number="3.5.5.7"		
	/codon_start=1		
	/evidence=experimental		
gene	/product="nitrilase"		
	/protein_id="AA970385.1"		
	/db_xref="GI:40890081"		
	/translation="MLSPTDGYRAAVOAPSFLDLDTREKTAIIIEQAADQVRLLI APEWTVIGYPLMIWLGSFAMGMPFVGCFENSLVRSGKNWNALADAARHRMTVVVGF FSRFGGSLVMGOALFGPEGELIARRRLKETHARETVFGSDSHLAVETGVGRIG		
CDS			

Query Match	30.7%	Score 311.4	DB 3	Length 1026
Best Local Similarity	59.0%	Pred. No. 4.3e-81		
Matches 534	Conservative 0	Mismatches 371	Indels 0	Gaps 0
QY	19	GTGCGCTGCGGCAAGCCGCGCCCATCTCACTGATTTGGAGGGGACGGGTGACAAACC	78	
Db	28	GCGCGCGGGGTGCGAGCGCGGCCCATCTTTCTCGATCTCGACCGCACCTGTGAGAGACG	87	
QY	79	ATTGAGTTGATGAGAAAGACGACGTAATATGCTCGTCTGATCGCTCTTCCGAAACT	138	
Db	88	ATTCGGATCATCGAGACAGCGGGCCGAGAGATGTGGCCGTGATCGGCTTCCGAAACC	147	
QY	139	TGGAATTCAGGCTACCATGGTTTCTTTGGCTTGACTCACCAGCATGGGCAATGCATTT	198	
Db	148	TGGAATCCCGGCTATCCGCTCTGGAATCGGCTCGGCTCGCGGGCGATCGGCTTC	207	
QY	199	GTAAGCCAAATACATAGAACTCATGTGAGTTGATGGCCCTCAAGCTAACGCAATTTCA	258	
Db	208	GTGACGCGCTTATTCGAAACTCTCCCTGGTGGCGGACGAAACGTGGAAACGCAATCGCC	267	
QY	259	GATGACAGCAACGGTGTGGGAATGATGGTTCACCTCGGGATGAGTGAACGGGTGCTGGC	318	
Db	268	GATGCGGGGGGGCCGACCGCATGACGTCGTCTGTGGCTTCACGACGAGCGGCGGAGGC	327	
QY	319	ACCCCTTAACATCAGTCAGTGGTTGATAGGCGATATAGTGAACACATTTGGGGCCCGGCA	378	
Db	328	AGCCTCTACATGGGCGACAGCGATCTTCGGCCCCGAAAGCGAGCTCATGCGGGCGCGCG	387	
QY	379	AAGTTGAACCTACTTCTTTGTGAACGTACTTTTGGCGGAAGGGGATGTTTCAATCGTA	438	
Db	388	AAAGCTCAAGCCGACACACGCGGACGAAACGGTGTGGCGGCGGACGAGCCCATCTTG	447	
QY	439	GCGGTTTTCGAGACGCTGTGTGAAGAGCGTGGGTGGCTTATGCTGTGGGACACCTTCAA	498	
Db	448	GCCGTTTACGAGACGGGGTGTGTGCGATCGGCGCCCTGTCTGTGGAGCAATCCAG	507	
QY	499	CCGCTTAACAAATACGCTTTGTATGCAACAAATGAAGATTCATTGTGCGGCTTGGCGG	558	
Db	508	CCGCTCTGAAATACGCGATGTATGCGGCCAACGAACGAGTGCAATGTCGCTCGTGGCGG	567	
QY	559	AGCTTTAGCCTTTATCTCAATGCGGCGGAAGCCCTGGGGCCTGATGTCAATGTAGCGGC	618	
Db	568	TGCTTCACCTTTTATTCGGGCGCATGGCTTATGCGCTCGGCGGAGGTGAACACGCGCGG	627	
QY	619	TCTGGAATCTATGCGCGTTGAAGGGGCATAGCTGTGATCACTAGCGGTGCGGCTCGTTCA	678	
Db	628	AGCCAGGCTTACGCGGTGCGAGGGCGGCTGCTACGTGCTGGCCTCTGTCTGTGTGTACA	687	
QY	679	CAATTCATGATCGATATGCTTTGTACAGATGACGAAGAACAATGCGTCTTCTGCGTGT	738	
Db	688	CCCGAGATCTGAAGAGGTGCTGATGACACGCGCGAACGAAGAGCCGTTGTCTGCGCGGC	747	
QY	739	GATGGAACCTACGATATATAGGCGCTGTATGTGTGTGACTTGTGTGCGGCTCTTGGCGAA	798	
Db	748	GGGGGGTTTCTGATGATCTTTCGGCCCGGACGCGCGCGCTCGGCCACGCGCTCCGAG	807	
QY	799	AATGAAGGGGTATTTCTTACGAAACCTTGATCTGTGAAGTACGCACTTGTGCTAAATATG	858	
Db	808	ACCGAAGGGGGCTGTACGCGCGAGATCGATCTGGCGGATGCGGCTGCGCAAGGCC	867	
QY	859	GCGGACAGCCCTGTGCTTATTTCCGCTCCGACATTACTCGCTTGTCTAATAGATGCG	918	
Db	868	GCGGCGCATCCCGCGGCGATTATAGCGGCGGCGGACGATGACGGGTTGTGTGAACCGG	927	
QY	919	AGCCC 923		
Db	928	CGCCC 932		

RESULT 12  
AY487522 1077 bp DNA linear ENV 05-APR-2004  
LOCUS AY487522  
DEFINITION Uncultured organism clone 2A17 nitrilase (BD5263) gene, complete  
cde.  
ACCESSION AY487522  
VERSION AY487522.1 GI:40890248  
KEYWORDS ENV.  
SOURCE uncultured organism  
ORGANISM uncultured organism  
REFERENCE 1 (bases 1 to 1077)  
AUTHORS Robertson, D.E., Chaplin, J.A., Desantis, G., Podar, M., Madden, M.,  
Chi, E., Richardson, T., Milan, A., Miller, M., Weiner, D.P., Wong, K.,  
McQuaid, J., Farwell, B., Preston, L.A., Tan, X., Snead, M.A.,  
Keller, M., Mathur, E., Kretz, P.L., Burk, M.J. and Short, J.M.  
Exploring nitrilase sequence space for enantioselective catalysis  
Appl. Environ. Microbiol. 70 (4), 2429-2436 (2004)  
2 (bases 1 to 1077)  
AUTHORS Robertson, D.E., Chaplin, J.A., Desantis, G., Podar, M., Madden, M.,  
Chi, E., Richardson, T.H., Milan, A., Miller, M., Weiner, D.P., Wong, K.,  
McQuaid, J., Farwell, B., Preston, L.A., Tan, X., Keller, M., Mathur, E.,  
Kretz, P.L., Burk, M.J. and Short, J.M.  
Direct Submission  
Submitted (26-NOV-2003) Bioinformatics, Diversea Corporation, 4955  
Directors Place, San Diego, CA 92121, USA  
FEATURES  
source  
1..1077  
/organism="uncultured organism"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:155900"  
/clone="2A17"  
/environmental\_sample  
1..1077  
/locus\_tag="BD5263"  
1..1077  
/locus\_tag="BD5263"  
/EC\_number="3.5.5.7"  
/codon\_start=1  
/evidence="experimental"  
/product="nitrilase"  
/protein\_id="AAR97469.1"  
/db\_xref="GI:40890249"  
/translation="MALTRPKLKVAAVQAPAPLDVDAVYKAVRLIDEAANGSLV  
APETWIPGFPMWLSPPAWALMRFVSRKFDNSLSDSRQAEHLRDAARHKLTIV  
MGLSERAGSLSYIAQWIIIGPNEGIVQARKLPTAERTVFGEGSHLAVNLPVIGR  
LGALCWEHLQPLSKYAMQNEIIVHAWPESFLYDPAFALGAEVNNASQIYAVE  
GSCFVAPCAVISOEMIDMCDTBEKHQHLIVHGGFTVIYGPDGRILDKLAPDEGI  
VYADIDGMIPILAKAADPACHYARPDTRLLFNNRPANRVETVLIPVDQVRDIDARV  
EAAPQARPATGNEDPAKPMMAE"  
ORIGIN  
Query Match 30.5%; Score 309.4; DB 3; Length 1077;  
Best Local Similarity 59.8%; Pred. No. 1.7e-80;  
Matches 539; Conservative 0; Mismatches 356; Indels 6; Gaps 1;  
QY 14 TCAAGTGGCTGCTGCAAGCCGCCGATTAATGATTTGAGAGCGGATGAC 73  
DB 23 TGAAGTGGCCGCGCGTGAAGGCGAGCGCTTCCTGATTCGATGCGGACATG 82  
QY 74 AAACATTTGAGTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 133  
DB 83 AAGCGTGGCGTTAATCGACGAGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAG 142  
QY 134 AAACATTTGAGTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 191  
DB 143 AGACCTGGATCCCGGCTATCCGTTTGGATCTGGCTGGCTGGCGGCTGGCAATCA 202  
QY 192 ----GCAATTTTGAAGCCCAATCACTAGAACTCATTTGAGAGTTGAGTGCCTCAAGCTA 247  
DB 203 TGGCGGGGTGGTGTCTGGCTATTTTCGATTAATTCGCTCAGCTATGACAGCCGCGAGGAG 262

QY 248 AGCGCATTTGAGATGACGAGCGGTTGGGAATCATGTCACCTTGGGATGAGTAC 307  
DB 263 AGCGCTGGCGGAGCGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 322  
QY 308 GGGTCGGTGGACACCTTTATCATGATGATGATGATGATGATGATGATGATGATG 367  
DB 323 GCGCGGCGGAGTGGCTTTATCATGATGATGATGATGATGATGATGATGATGATG 382  
QY 368 GGGCGCGGAG 427  
DB 383 CACAGCGGCGGAG 442  
QY 428 GTTCATGCTGAGCGGTTTGGAGAGCTGTGTGGAAGGCTGGGAGCTTATGCTGTTGG 487  
DB 443 GCAGCCACCTGGCGGATACATATTTCAATCGACGCGCTGCGCTGCTGCTGCGG 502  
QY 488 AGCACCTTCAACCGCTTAACAAATATGATGATGATGATGATGATGATGATGATG 547  
DB 503 AGCACCTTCAACCGCTTCAACAAATATGATGATGATGATGATGATGATGATGATG 562  
QY 548 CGGCTTGGCGGAGCTTTAGCTTTATCTTATGATGATGATGATGATGATGATGATG 607  
DB 563 CGGATGCGCGTCT 622  
QY 608 ATGTAGGCGGCTCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 667  
DB 623 ACAAGCAG 682  
QY 668 CGCTCGTTTCAATTCATGATGATGATGATGATGATGATGATGATGATGATGATGATG 727  
DB 683 CGGTGATCTCGACAGAAATGATGATGATGATGATGATGATGATGATGATGATGATGATG 742  
QY 728 TTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 787  
DB 743 TTTCAGTGGTGGGCTTCAACGCTGATCTATGAGCGGAGCGGCGGATCGGCA 802  
QY 788 CTCTTGGCGAAATGAG 847  
DB 803 AGCTCGCGCGAGATCAG 862  
QY 848 TTGCTAAATGCGCGAG 907  
DB 863 TCGGAAAGTGGCGCGATCTCTCGGCACTATGCGAGCCGAGATTACCGCCTTC 922  
QY 908 T 908  
DB 923 T 923  
RESULT 13  
LOCUS AY487430  
DEFINITION Uncultured organism clone 2A15 nitrilase (BD538) gene, complete  
cde.  
ACCESSION AY487430  
VERSION AY487430.1 GI:40890064  
KEYWORDS ENV.  
SOURCE uncultured organism  
ORGANISM uncultured organism  
REFERENCE 1 (bases 1 to 1047)  
AUTHORS Robertson, D.E., Chaplin, J.A., Desantis, G., Podar, M., Madden, M.,  
Chi, E., Richardson, T., Milan, A., Miller, M., Weiner, D.P., Wong, K.,  
McQuaid, J., Farwell, B., Preston, L.A., Tan, X., Snead, M.A.,  
Keller, M., Mathur, E., Kretz, P.L., Burk, M.J. and Short, J.M.  
Exploring nitrilase sequence space for enantioselective catalysis  
Appl. Environ. Microbiol. 70 (4), 2429-2436 (2004)  
2 (bases 1 to 1047)  
AUTHORS Robertson, D.E., Chaplin, J.A., Desantis, G., Podar, M., Madden, M.,  
Chi, E., Richardson, T.H., Milan, A., Miller, M., Weiner, D.P., Wong, K.,  
McQuaid, J., Farwell, B., Preston, L.A., Tan, X., Keller, M., Mathur, E.,  
Kretz, P.L., Burk, M.J. and Short, J.M.

TITLE Direct Submission  
JOURNAL Submitted (26-NOV-2003) Bioinformatics, Diversa Corporation, 4955  
Directors Place, San Diego, CA 92121, USA  
FEATURES  
source Location/Qualifiers  
1..1047  
/organism="uncultured organism"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:155900"  
/clone="2A15"  
/environmental\_sample  
1..1047  
/locus\_tag="BD5338"  
1..1047  
/locus\_tag="BD5338"  
/EC\_number="3.5.5.7"  
/codon\_start=1  
/evidence="experimental"  
/product="nitrilase"  
/protein\_id="AAR97377.1"  
/db\_xref="GI:40890065"  
/translation="MPTSKQPRVAAVQAAPVFLDEGALISKGISIEEASNGAKLIA  
FPEWTWPGYPMWILDSPAMGMRFVQRYFDSLMIGSQARMMQAAANNKIYVMGY  
SERSGSLYMGSIIINDGETIFTRKXLPHTVERTVEBDGSHLCYMDTEIGVGA  
MCWEHLQPLSKYAMYSDOEIOHIASMPSFLYGAAYALGPELNNASOMYAAEGQC  
FLAPCATVSKEMIEMLIDDPKKEPILLEGGFMTIYGPDRPLAKPLPENEBGLIYA  
DIDGKMSMARAADPRAGHYARPDVTRLLFNASAPNRVEYINPASGPRESIKDMQKMQ  
MEAEQKALREMT"

## ORIGIN

Query Match 30.4%; Score 308.4; DB 3; Length 1047;  
Best Local Similarity 58.9%; Pred. No. 3.4e-80;  
Matches 531; Conservative 0; Mismatches 371; Indels 0; Gaps 0;

14 TCAGAGTGCCTGCGTGAAGCCGCCGATCTACATGATTTGGAGGCGACGTGACA 73  
20 TTAGAGTGCCTGCGTGAAGCCGCCGATCTACATGATTTGGAGGCGACGTGACA 79  
74 AAACATGATGATGATGAGAAAGACGACCAATATATGCTGCTGATCCGCTTCCGG 133  
80 AAGGATCTCCCTCTTATGAGAGGCGGCTCCCAATGAGCCAAAGCTCATGCTTCCGG 139  
134 AAACATGATGATGATGAGAAAGACGACCAATATATGCTGCTGATCCGCTTCCGG 193  
140 AAACATGATGATGATGAGAAAGACGACCAATATATGCTGCTGATCCGCTTCCGG 199  
194 AATTTGATGCCAATATACATGAGAACTGATGAGATTGAGTGGCCCTCAAGCTAAGCC 253  
200 GCTTGTCAAGGCTATTTTGAACAATGCTGATGCTGGGTATGAGCAAGCCAAAGCC 259  
254 TTTGATGATGAGAAAGACGACCAATATATGCTGATGCTGGGTATGAGCAAGCC 313  
260 TGAACAGAGCTGCGCCCAATATACAAAGATTTAGTGTATGAGCGCAATCCAGTG 319  
314 GTGAGACCTTTATCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 373  
320 GTGAGACCTTTATCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 379  
374 GTGAGACCTTTATCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 433  
380 GTGAGACCTTTATCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 439  
434 GTGAGACCTTTATCATGATGATGATGATGATGATGATGATGATGATGATGATGAT 493  
440 ATCTTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 499  
494 TTCAACCGCTAACAATATGATGATGATGATGATGATGATGATGATGATGATGATGAT 553  
500 TGCACCGCTTACGCAATATGCAATGATTTCTCAAGATGAAACAATTTCAATGCTCT 559  
554 GGCACGCTTTAGCTTTATCTTAATGCGGCGAAGCCCTGGGCGCTGATGCTCAATGTA 613  
560 GGCACGCTTTAGCTTTATCTTAATGCGGCGAAGCCCTGGGCGCTGATGCTCAATGTA 619

QY 614 CGGCTCTCGAATCTATGCGCTTGAAGGCGCAATGCTTGTATCTAGCTGCTGCGCTCG 673  
DB 620 CGGCGACCAAAATGATGAGCGCGGACGAGCTGCTTGTGCTTGTGCTTGTGCTGCGCA 679  
QY 674 TTTCACAAATCCATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 733  
DB 680 TCTCAAAAGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 739  
QY 734 CTGCTGCTGAGCACTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 793  
DB 740 AAGGTGGCGGTTACCATGATTTTACCGCCCGATGGCGACCTTGTGCTTAAACGTTTC 799  
QY 794 CCGAAATGAAAGAGGATTTCTTACCGCAAACTTGTATCTTGAAGTATGATGATGATGAT 853  
DB 800 CTGAAAGAGGAGGCGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 859  
QY 854 AATATGCGGACGACCTGCTGCTGATTTATTTCCGCTCCGACATTACTGCTTGTATGAT 913  
DB 860 AGGCTGCGCGGACCGCGGACGATGATGATGATGATGATGATGATGATGATGATGATGAT 919  
QY 914 AT 915  
DB 920 AT 921  
RESULT 14  
AY487543 1062 bp DNA linear ENV 05-APR-2004  
LOCUS Uncultured organism clone 2A20 nitrilase (BD7753) gene, complete  
DEFINITION cds.  
ACCESSION AY487543  
VERSION AY487543.1 GI:40890290  
KEYWORDS ENV.  
SOURCE uncultured organism  
ORGANISM uncultured organism  
AUTHORS unclassified; environmental samples.  
1 (bases 1 to 1062)  
REFERENCE  
1. Robertson, D.E., Chaplin, J.A., Desantis, G., Podar, M., Madden, M.,  
Chi, E., Richardson, T., Milan, A., Miller, M., Weiner, D.P., Wong, K.,  
McQuaid, J., Farwell, B., Preston, J.A., Tan, X., Sneed, M.A.,  
Keller, M., Mathur, E., Kretz, P.L., Burk, M.J. and Short, J.M.  
Exploring nitrilase sequence space for enantioselective catalysis  
Appl. Environ. Microbiol. 70 (4), 2429-2436 (2004)  
2 (bases 1 to 1062)  
AUTHORS  
1. Robertson, D.E., Chaplin, J.A., Desantis, G., Podar, M., Madden, M.,  
Chi, E., Richardson, T.H., Milan, A., Miller, M., Weiner, D.P., Wong, K.,  
McQuaid, J., Farwell, B., Preston, J.A., Tan, X., Keller, M., Mathur, E.,  
Kretz, P.L., Burk, M.J. and Short, J.M.  
TITLE Direct Submission  
JOURNAL Submitted (26-NOV-2003) Bioinformatics, Diversa Corporation, 4955  
Directors Place, San Diego, CA 92121, USA  
FEATURES  
source Location/Qualifiers  
1..1062  
/organism="uncultured organism"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:155900"  
/clone="2A20"  
/environmental\_sample  
1..1062  
/locus\_tag="BD7753"  
1..1062  
/locus\_tag="BD7753"  
/EC\_number="3.5.5.7"  
/codon\_start=1  
/evidence="experimental"  
/product="nitrilase"  
/protein\_id="AAR97490.1"  
/db\_xref="GI:40890291"  
/translation="MGVAHPKTYKIAVAQAPAFIDLDASVEKAVRFIDEAGAGARLI  
AFPEWTWPGYPMWILGAPAWALIRGVSRYFDSLSYDSQAKRLAAARNNRVV  
LGSEBDGSLYIAQWITGPDETIARRKXLPHTARTVEBDGSHLAHYEDVGR  
LGALCWEHLQPLSKYAMQNEQVHAAMPSPFLYDPFAHALCAEVNNAASKIYAVE  
GSCFVIAPCATVSGAMIDELCDTPBEKIQFLHAGGFANIVGPDGAPLAAPLPDKEGL



## ORIGIN

LYADIDGMSIVAKAADBPAGHYARPDVTLFPNNRBYRVETMALPIDETKAEARA  
KPEKAPNVAFAPVQAAE"

Query Match 30.3%; Score 307; DB 3; Length 1062;  
Best Local Similarity 58.0%; Pred. No. 8.9e-80;  
Matches 548; Conservative 0; Mismatches 375; Indels 6; Gaps 1;

QY 15 CAAGATGCGCTGCGTGAAGCCGCCGATCTACATGATTTGAGCGGACGATGACAA 74  
DB 24 CAAGTGGCCCGCGGACGAGCGCCGCTTTTCTCACTGACCGCTCGTGGCAAAA 83  
QY 75 AACCATTAAGTTAGTGAAGAAGACGACGATTAATGCTGTCTGATCGCTTCCGGA 134  
DB 84 GGGCGTCCGTTTCATCGAAGCGCGCGCGCGCGCGCGCGCTCATGCGCTTCCGGA 143  
QY 135 AACTTGAATTCAGGCTACCATGCTTCTTTGGCTTACTACACGAGATGGCAAT--- 191  
DB 144 GACCTGATACCCGGTTACCCCTGTGATCTGGCTAGCGCGCGCTGGGCTATCAT 203  
QY 192 ---GCAATTTGATGCCCAATACATGAGAACTCATTTGATGGCTCAAGCTAA 248  
DB 204 GCGCGCTTCGCTTCGCGCTATTTTCAGACACTCGCTCAGTACGACAGCCGCGCA 263  
QY 249 GCGCATTTCAATGAGCGACAGCGGTTGGGAATATGTCACCCCTGGGGATGATGAACG 308  
DB 264 GAACTCCGCGCGCGCGCAAGCGCAACAAGTGGTGGTGTCTCGGCTCTCCGACG 323  
QY 309 GGTGCGGACACCTTTTACATCAGTCACTGCTTACGAGGATTAATGATGACACCATTTG 368  
DB 324 CGAGCGGCGGACGCTTTTACATGCGCATGATATCGGCCGCGGACGCGCAACATGCG 383  
QY 369 GGGCGGCGAAAGTTGAACCTACTTTTGTGAACGTACTTTTGGCGGAAAGGGAATG 428  
DB 384 CAAGCGCGCGCAAGCTCAAGCGCACGCGAGGCGGACGCTTCCGCGAAAGGGAAG 443  
QY 429 TTCAATGCTAGCGGTTTTCAGAGAGCTGTGTTGAAGCGGCGGCTTATGCTGTGGA 488  
DB 444 CTCGATCTTGGGAGTGCAGAGCTGATGTTGGCGGCTCGGCGGCTGTGCTGTGGA 503  
QY 489 GCACCTTCAACCGCTTACAAATAATACGCTTTGATGCAAAAATGAAGATTCATTTGCG 548  
DB 504 ACACCTGACGCGGCTGTCAAAATACGCGCATGATGCGAGAAAGAGTGCATGTCGC 563  
QY 549 GCGTTGGCGGAGCTTTAGCTTTATCTTAATGCGCGGAAAGCGCTGGGCGCTGATGCA 608  
DB 564 GCGCTGGCGGAGCTTTTGGCTTTAGATCCGTTCCGCAAGCGGCTGGCGGGAAGTAA 623  
QY 609 TGTAGCGGCTCTCGAATCTATGCGTTGAAGGGAATGCTTGTACTAGCGTGTGTCG 668  
DB 624 CATGCGGCGGAGCAAAATCTATGCGTGAAGGCTCGTGTGTCATGCGCGCTGCGC 683  
QY 669 GCTGCTTTCACAAATCATGATGATATGTTTGTACAGATGACGAAAGATGCGTTGCT 728  
DB 684 GACCGTTTTCGACGAGATGATGACGAACTCTGCAATACGCGGAGAGATCACTTCT 743  
QY 729 TCTGCTGTGTGTGACACTCACTGATATGATAGGCGCTGATGCTGATGCTGTCGCC 788  
DB 744 GATCCGCGGCGGCGCTTTCGCTGATTTTACGCGCCGACGCGCGCGCTGCGCGCGC 803  
QY 789 TCTTCCGCAAAATGAAGAGGATTTCTTACGCAAACTTGAATCTGAGATGCGATCCT 848  
DB 804 GCTGCGCGCGGACAAAGAGGCTTGTCTTACGCGCATGATCTTCGGGATATTTGGT 863  
QY 849 TGTGTAATGCGGAGACCTGCTGCTGATTAATTTCCGCTCCGACATTAATCTGCTTGT 908  
DB 864 TGCAGAAAGGCGAGCGATCCGCGCGGCGATTAATGAGCGCCGACGATCAACCGCTTCT 923  
QY 909 AATGATGCGAGCGCTTAATTTACCGGTAG 937  
DB 924 GTTCAACAATCGCGCTGGGTATCGGATCG 952

## RESULT 15

AY487497

LOCUS

AY487497 1041 bp DNA linear ENV 05-APR-2004

DEFINITION

Uncultured organism clone 2A13 nitrilase (BD5086) gene, complete cds.

ACCESSION

AY487497

VERSION

AY487497.1 GI:40890198

KEYWORDS

ENV.

SOURCE

uncultured organism

ORGANISM

uncultured organism

REFERENCE

1 (bases 1 to 1041)  
Robertson,D.E., Chaplin,J.A., Desantis,G., Podar,M., Madden,M.,  
McQuaid,J., Farwell,B., Preston,L.A., Tan,X., Sneed,M.A.,  
Keller,M., Mathur,E., Kretz,P.L., Burk,M.J., and Short,J.M.  
Exploring nitrilase sequence space for enantioselective catalysis  
Appl. Environ. Microbiol. 70 (4), 2429-2436 (2004)

AUTHORS

Robertson,D.E., Chaplin,J.A., Desantis,G., Podar,M., Madden,M.,  
Chi,E., Richardson,T.H., Milan,A., Miller,M., Weiner,D.P., Wong,K.,  
McQuaid,J., Farwell,B., Preston,L.A., Tan,X., Keller,M., Mathur,E.,  
Kretz,P.L., Burk,M.J., and Short,J.M.

TITLE

Direct Submission

JOURNAL

Submitted (26-NOV-2003) Bioinformatics, Diversa Corporation, 4955  
Directors Place, San Diego, CA 92121, USA

FEATURES

Location/Qualifiers

SOURCE

1..1041  
/organism="uncultured organism"  
/mol\_type="genomic DNA"  
/db\_xref="taxon:155900"  
/clone="2A13"  
/environmental\_sample  
1..1041  
/locus\_tag="BD5086"  
/locus\_tag="BD5086"  
/BC\_number="3.5.5.7"  
/evidence=experimental  
/product="nitrilase"  
/protein\_id="AAR9444.1"  
/db\_xref="GI:40890199"  
/translation="MSEPMTRYGAOVAQAVFLLDRTVEKAIGLIEQAQDVRLL  
APERTWIPGVPMIWLGAIPAMGRFVQREYENSIVRSKOMOLADARHGMVVG  
YSEBAGSLVMGQAIRPDDDLTAARPKKPTAERVPREGSHLAVDTAIGRG  
ALCCWEHIQPLSKYAMTADEQYHVAWSFSLYRGAVALGPEVNTAAQIYAVEG  
CYVLASCATVSPENIKVLDTPDKEMFLKAGGGFAMFGPDGRALAEPLPETERGLV  
ADIDIGMIALAKMAADBPAGHYSRPDVTRLIDRRPAQVVTLDAAFEPQREKGDABA  
LRVAVESAAAAQ"

gene

1..1041  
/locus\_tag="BD5086"  
/locus\_tag="BD5086"  
/BC\_number="3.5.5.7"  
/evidence=experimental  
/product="nitrilase"  
/protein\_id="AAR9444.1"  
/db\_xref="GI:40890199"  
/translation="MSEPMTRYGAOVAQAVFLLDRTVEKAIGLIEQAQDVRLL  
APERTWIPGVPMIWLGAIPAMGRFVQREYENSIVRSKOMOLADARHGMVVG  
YSEBAGSLVMGQAIRPDDDLTAARPKKPTAERVPREGSHLAVDTAIGRG  
ALCCWEHIQPLSKYAMTADEQYHVAWSFSLYRGAVALGPEVNTAAQIYAVEG  
CYVLASCATVSPENIKVLDTPDKEMFLKAGGGFAMFGPDGRALAEPLPETERGLV  
ADIDIGMIALAKMAADBPAGHYSRPDVTRLIDRRPAQVVTLDAAFEPQREKGDABA  
LRVAVESAAAAQ"

CDs

1..1041  
/locus\_tag="BD5086"  
/locus\_tag="BD5086"  
/BC\_number="3.5.5.7"  
/evidence=experimental  
/product="nitrilase"  
/protein\_id="AAR9444.1"  
/db\_xref="GI:40890199"  
/translation="MSEPMTRYGAOVAQAVFLLDRTVEKAIGLIEQAQDVRLL  
APERTWIPGVPMIWLGAIPAMGRFVQREYENSIVRSKOMOLADARHGMVVG  
YSEBAGSLVMGQAIRPDDDLTAARPKKPTAERVPREGSHLAVDTAIGRG  
ALCCWEHIQPLSKYAMTADEQYHVAWSFSLYRGAVALGPEVNTAAQIYAVEG  
CYVLASCATVSPENIKVLDTPDKEMFLKAGGGFAMFGPDGRALAEPLPETERGLV  
ADIDIGMIALAKMAADBPAGHYSRPDVTRLIDRRPAQVVTLDAAFEPQREKGDABA  
LRVAVESAAAAQ"

ORIGIN

Query Match 30.1%; Score 305.4; DB 3; Length 1041;  
Best Local Similarity 58.0%; Pred. No. 2.7e-79;  
Matches 540; Conservative 0; Mismatches 391; Indels 0; Gaps 0;

QY 19 GTGCGTGGCGGCAAGCGCGCCGATCTACATGATTTGAGCGGACGATGACAAACC 78  
DB 28 GGGCGGCGGAGTGCAGCGCGCGCGCTGCTTCATCTCGACCGCAAGTGAAGAAACG 87  
QY 79 ATTGATGATGAGGAAGACGACGATTAATGCTGCTGATGCTGCTTTCCGAAACT 138  
DB 88 ATCGGCTGATCGAGCGAGCGGCGCAAGAGACGTGGCGCTGATCGATTTCCAGAGACT 147  
QY 139 TGAATTCAGGCTTACCATGCTTTCTTTGGCTTGAATCTACAGCATGGGCAATGCAATT 198  
DB 148 TGAATTCGCGCTTATCCCTTTTGGATATGCTGCGCGCGCGCTTGGGCGCATCGCTTC 207  
QY 199 GTAAGCCAAATCAATGAACTCATGAGATGATGAGGCGCTCAAGCTAAGCATTTCA 258  
DB 208 GTTCAAGCGCTTATTTGAGAAATTCGCTGCTGCGCGGACCAAGCATGAGGAGCGCTGCG 267  
QY 259 GATGAGCAAGCGGCTTGGGAATCATGATGATCACTGAGGAGATGATGAACGCGGTGCTGCG 318





**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 26, 2006, 22:18:14 ; Search time 701.165 Seconds  
(without alignment)  
9638.243 Million cell updates/sec

Title: US-09-751-299-3  
Perfect score: 1014  
Sequence: 1 atgaagaagcgtatcacaagt.....cgcaactcgagaatttga 1014

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 4996997 seqs, 3332346308 residues

Total number of hits satisfying chosen parameters: 9993994

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : N\_Geneseq\_21:\*

- 1: geneseqn1980s:\*
- 2: geneseqn1990s:\*
- 3: geneseqn2000s:\*
- 4: geneseqn2001s:\*
- 5: geneseqn2002as:\*
- 6: geneseqn2002bs:\*
- 7: geneseqn2003as:\*
- 8: geneseqn2003bs:\*
- 9: geneseqn2003cs:\*
- 10: geneseqn2003ds:\*
- 11: geneseqn2004as:\*
- 12: geneseqn2004bs:\*
- 13: geneseqn2004cs:\*
- 14: geneseqn2005s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1014	100.0	1014	4	AD1167 Nitriase
2	1012.4	99.8	1014	10	AD1167 Nitriase
3	1012.4	99.8	1014	12	AD1167 Nitriase
4	1012.4	99.8	1014	12	AD1167 Nitriase
5	1012.4	99.8	1014	12	AD1167 Nitriase
6	1012.4	99.8	1014	12	AD1167 Nitriase
7	1012.4	99.8	1014	12	AD1167 Nitriase
8	1012.4	99.8	1014	12	AD1167 Nitriase
9	1012.4	99.8	1014	12	AD1167 Nitriase
10	1012.4	99.8	1014	12	AD1167 Nitriase
11	1012.4	99.8	1014	12	AD1167 Nitriase
12	1012.4	99.8	1014	12	AD1167 Nitriase
13	1012.4	99.8	1014	12	AD1167 Nitriase
14	1012.4	99.8	1014	12	AD1167 Nitriase
15	1012.4	99.8	1014	12	AD1167 Nitriase
16	1012.4	99.8	1014	12	AD1167 Nitriase
17	1012.4	99.8	1014	12	AD1167 Nitriase
18	1012.4	99.8	1014	12	AD1167 Nitriase
19	1012.4	99.8	1014	12	AD1167 Nitriase

20	399.6	39.4	999	12	AD162271	Adi62271 DNA encod
21	399.6	39.4	999	12	AD164392	Adi64392 DNA encod
22	396.4	39.1	1002	10	AD164024	Adi64024 DNA encod
23	396.4	39.1	1002	12	AD164125	Adi64125 Chemical
24	396.4	39.1	1002	12	AD164223	Adi64223 DNA encod
25	396.4	39.1	1002	12	AD164223	Adi64223 DNA encod
26	396.4	39.1	1002	12	AD164544	Adi64544 DNA encod
27	364.2	35.9	1038	10	AD164066	Adi64066 DNA encod
28	364.2	35.9	1038	12	AD164167	Adi64167 Chemical
29	364.2	35.9	1038	12	AD164265	Adi64265 DNA encod
30	364.2	35.9	1038	12	AD164586	Adi64586 DNA encod
31	358.6	35.4	1068	10	AD164038	Adi64038 DNA encod
32	358.6	35.4	1068	12	AD164139	Adi64139 Chemical
33	358.6	35.4	1068	12	AD164247	Adi64247 DNA encod
34	358.6	35.4	1068	12	AD164558	Adi64558 DNA encod
35	358.6	35.4	1068	12	AD164558	Adi64558 DNA encod
36	319	31.5	1011	10	AD164051	Adi64051 Chemical
37	319	31.5	1011	12	AD164051	Adi64051 Chemical
38	319	31.5	1011	12	AD164051	Adi64051 Chemical
39	319	31.5	1011	12	AD164051	Adi64051 Chemical
40	319	31.5	1011	12	AD164051	Adi64051 Chemical
41	313.8	30.9	1080	8	AD164384	Adi64384 Chemical
42	313.8	30.9	1080	8	AD164384	Adi64384 Chemical
43	313.8	30.9	1080	8	AD164384	Adi64384 Chemical
44	313.8	30.9	1080	8	AD164384	Adi64384 Chemical
45	311.4	30.7	1026	10	AD164384	Adi64384 Chemical

## ALIGNMENTS

RESULT 1	
ID	AD1167 standard; DNA; 1014 BP.
AD1167	
AC	AD1167
DT	25-SEP-2001 (first entry)
DE	Nitriase DNA #2.
XX	Nitriase; enantiomer; alpha-substituted carboxylic acid;
XX	stereoselective; hydrolysis; amino nitrite; cyanohydrin; ds.
OS	Unidentified.
FT	Key
FT	CDS
FT	Location/Qualifiers
FT	1..1041
FT	/*tag=a
FT	/product="Nitriase protein #2"
XX	
XX	MO200148175-A2.
XX	
XX	05-JUL-2001.
XX	
XX	29-DEC-2000; 2000MO-US035555.
XX	
XX	29-DEC-1999; 99US-0173609P.
XX	
XX	07-DEC-2000; 2000US-0254414P.
XX	
XX	(DIVE-) DIVERSA CORP.
XX	(MADP/) MADP D.
XX	
XX	Madden M, Weiner DP, Chaplin JA;
XX	
XX	WPI; 2001-465211/50.
XX	
XX	P-PSDB; AAE05489.
XX	
XX	Producing alpha-substituted carboxylic acid enantiomers by contacting
XX	aldehyde or ketone with cyanide, ammonia compound or its salt or amine
XX	and hydrolyzing the resulting amino nitrite or cyanohydrin with
XX	Nitriase.

PS Claim 28; Page 85-86; 87pp; English.

CC The present invention relates to methods for producing an  
CC enantiomerically pure alpha-substituted carboxylic acid. The method  
CC involves contacting an aldehyde or ketone with a cyanide containing  
CC compound and an ammonia-containing compound or an ammonium salt or amine,  
CC and stereoselectively hydrolyzing the resulting amino nitrile or  
CC cyanohydrin intermediate with a nitrilase or a polypeptide having  
CC nitrilase activity. The present sequence is a DNA encoding nitrilase #2  
XX  
XX Sequence 1014 BP; 237 A; 232 C; 280 G; 265 T; 0 U; 0 Other;

Query Match 100.0%; Score 1014; DB 4; Length 1014;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1014; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

QY 1 ATGAAAGAGCTATCAAGAGTGGCGGCGTGAAGCGCGCCGATCATGAGATTGGAG 60
DB 1 ATGAAAGAGCTATCAAGAGTGGCGGCGTGAAGCGCGCCGATCATGAGATTGGAG 60
QY 61 GCGACGCTGACAAAGCATTGATGATGAGAGAGAGCAGTATATGCTGCTG 120
DB 61 GCGACGCTGACAAAGCATTGATGATGAGAGAGAGCAGTATATGCTGCTG 120
QY 121 ATGCGCTTCCGGAAGCTTGATTCAGGCTTACCATGCTTCTTGGCTTGACTGACCA 180
DB 121 ATGCGCTTCCGGAAGCTTGATTCAGGCTTACCATGCTTCTTGGCTTGACTGACCA 180
QY 181 GCATGGGAGATGCAATTTGTATGAGCAATACATGAGAACTCATTTGAGTTGGATGGCCCT 240
DB 181 GCATGGGAGATGCAATTTGTATGAGCAATACATGAGAACTCATTTGAGTTGGATGGCCCT 240
QY 241 CAAGCTAAGCGGATTCAGATGACAGCGGCTTGGAGATATGCTGACCTGGGGATG 300
DB 241 CAAGCTAAGCGGATTCAGATGACAGCGGCTTGGAGATATGCTGACCTGGGGATG 300
QY 301 ACTGAACGGGTGGGTGGGACCCCTTATCATGATGATGCTTATGAGGAGATATGCTGAC 360
DB 301 ACTGAACGGGTGGGTGGGACCCCTTATCATGATGATGCTTATGAGGAGATATGCTGAC 360
QY 361 ACCATTGGGGCCCGGAGAAAGTTGAAGCTTATTTGTTGAACGTAATTTGTTGGCGAA 420
DB 361 ACCATTGGGGCCCGGAGAAAGTTGAAGCTTATTTGTTGAACGTAATTTGTTGGCGAA 420
QY 421 GGGGATGCTTATCGGTACCGGCTTTCAGAGCTCTGTTGGAAGGCTGGGTGCTTATG 480
DB 421 GGGGATGCTTATCGGTACCGGCTTTCAGAGCTCTGTTGGAAGGCTGGGTGCTTATG 480
QY 481 TGTGGGAGACCTTCAACCGCTAACAATAATGCTTGTATGCACAAAATGAAAGAT 540
DB 481 TGTGGGAGACCTTCAACCGCTAACAATAATGCTTGTATGCACAAAATGAAAGAT 540
QY 541 CATTTGCGGCTTTGCGGAGCTTTAGCTTTATCTTAATGCGGCGAAACCTTGGGCT 600
DB 541 CATTTGCGGCTTTGCGGAGCTTTAGCTTTATCTTAATGCGGCGAAACCTTGGGCT 600
QY 601 GATGTAATGTAAGCGGCTCTGAAATCTATGCGGTGAAGGCGAATGCTTCTGTAAGG 660
DB 601 GATGTAATGTAAGCGGCTCTGAAATCTATGCGGTGAAGGCGAATGCTTCTGTAAGG 660
QY 661 TCGTGTGCGCTGCTTCAACATTCATGATGATGCTTGTATGAGATGAGAAAGCAT 720
DB 661 TCGTGTGCGCTGCTTCAACATTCATGATGATGCTTGTATGAGATGAGAAAGCAT 720
QY 721 GCGTGTGCTTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 780
DB 721 GCGTGTGCTTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 780
QY 781 GTCGCGCTCTTTCGCGAAATGAAAGGATATCTTACGCAAACTTGAATCTGAGAT 840
DB 781 GTCGCGCTCTTTCGCGAAATGAAAGGATATCTTACGCAAACTTGAATCTGAGAT 840
QY 841 CGCATCTTGTGTAATAATGGCGAGACCTTGTGCTGCTGCTGCTGCTGCTGCTGCT 900

```

```

DB 841 CGCATCTTGTGTAATAATGGCGAGACCTTGTGCTGCTGCTGCTGCTGCTGCTGCT 900
QY 901 CGCTGCTATAGATGAGACCCCTAATTAACCGGTAGTTGAATTAAGGTATCTTCT 960
DB 901 CGCTGCTATAGATGAGACCCCTAATTAACCGGTAGTTGAATTAAGGTATCTTCT 960
QY 961 CTTTACGCTTTGGGTGAAGGCTTGAAGAGGCTGGCACTCGAAGAAATTGA 1014
DB 961 CTTTACGCTTTGGGTGAAGGCTTGAAGAGGCTGGCACTCGAAGAAATTGA 1014

```

# RESULT 2

ADC24118  
ID ADC24118 standard; DNA; 1014 BP.

AC ADC24118;  
DT 18-DEC-2003 (first entry)

DE DNA sequence (SeqID 385) encoding a nitrilase enzyme.

KW gene; ds; nitrilase; nitrile; cyanohydrin; ammonia; biocatalyst;  
KW enantiomer; chiral medicine.

OS Unidentified.

PN WO200300840-A2.

PD 03-JAN-2003.

PE 15-MAY-2002; 2002WO-US015983.

PF 21-JUN-2001; 2001US-0300189P.

PR 30-JUL-2001; 2001US-0309006P.

PR 22-JAN-2002; 2002US-0351336P.

PA (DIVE-) DIVERSA CORP.

PA (MADD/) MADDEN D.

PI Madden M, Desantis G, Chaplin JA, Weiner DP, Milan A, Chi E;

P1 Short JM, Burk M;

DR MPI; 2003-201417/19.

DR P-PSDB; ADC24119.

PT Novel nitrilase polypeptide, useful for making (R)- or (S)-ethyl-4-cyano-

PT 3-hydroxybutyric acid or (R)- or (S)-mandelic acid or (S)- or (R)-phenyl

PT lactic acid derivative and for producing pharmaceutical composition, and

PT food additive.

PS Claim 1; SEQ ID NO 385; 560pp; English.

XX This invention relates to nitrilases and the nucleic acids that encode  
XX these enzymes thereof. Specifically, it refers to polypeptides that  
XX exhibit nitrilase activity, i.e. the ability to directly hydrolyse  
XX nitriles or cyanohydrins into their corresponding carboxylic acids and  
XX ammonia. Nitrilases have commercial utility as biocatalysts for use in  
XX the synthesis of enantiomerically pure aromatic and aliphatic amino  
XX acids, as well as hydroxy acids, which are important for the development  
XX of chiral medicines. Furthermore, the present invention describes  
XX nitrilases, isolated from mesophilic microorganisms, that have improved  
XX activity and stability at increased pH and temperature. They are also  
XX inexpensive, efficient catalysts, have broad substrate specificity and  
XX are capable of chiral differentiation. This polynucleotide is a DNA  
XX sequence that encodes a nitrilase enzyme of the invention.

SEQ Sequence 1014 BP; 238 A; 232 C; 279 G; 265 T; 0 U; 0 Other;

Query Match 99.8%; Score 1012.4; DB 10; Length 1014;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 1013; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

```
QY 1 ATGAAAGAGCTATCAAGGTCGCTGCGTGAAGCCGCCGATCTACATGATTTGAG 60
DB 1 ATGAAAGAGCTATCAAGGTCGCTGCGTGAAGCCGCCGATCTACATGATTTGAG 60
QY 61 GGGACGGTGGACAAACCATTTGATGATGAAAGACACGTAATTAATGCTGCTG 120
DB 61 GGGACGGTGGACAAACCATTTGATGATGAAAGACACGTAATTAATGCTGCTG 120
QY 121 ATGCGCTTTCCGGAACCTTGATTCAGGCTACCCATGGTTTCTTTGGCTGACTGACA 180
DB 121 ATGCGCTTTCCGGAACCTTGATTCAGGCTACCCATGGTTTCTTTGGCTGACTGACA 180
QY 181 GCATGGGCAATGCATTTGTACGCCAATACATGAGACCTCATTTGATGATGAGCCCT 240
DB 181 GCATGGGCAATGCATTTGTACGCCAATACATGAGACCTCATTTGATGATGAGCCCT 240
QY 241 CAAGCTAAGCGCATTTCAATGACGCAAGCGGTTGGAAATCATGGTCACTCTGGGATG 300
DB 241 CAAGCTAAGCGCATTTCAATGACGCAAGCGGTTGGAAATCATGGTCACTCTGGGATG 300
QY 301 AGTGAACGGGTGGTGGCAACCTTTACATCAGTCAATGATGATAGCGGATTAATGCTGAC 360
DB 301 AGTGAACGGGTGGTGGCAACCTTTACATCAGTCAATGATGATAGCGGATTAATGCTGAC 360
QY 361 ACCATTGGGGCCCGCGGAAGTTGAAACCTACTTTTGTGAAAGTACTTTGTTGGCGGAA 420
DB 361 ACCATTGGGGCCCGCGGAAGTTGAAACCTACTTTTGTGAAAGTACTTTGTTGGCGGAA 420
QY 421 GGGGATGGTTTATCGCTACGCGGTTTTCGAGACGCTGTTGGAAGGCTGGGTGGCTTATGC 480
DB 421 GGGGATGGTTTATCGCTACGCGGTTTTCGAGACGCTGTTGGAAGGCTGGGTGGCTTATGC 480
QY 481 TGTGGGAGCACCTTCAACCGCTAACAAATACGCTTTATGAGCCAAATAGAGAT 540
DB 481 TGTGGGAGCACCTTCAACCGCTAACAAATACGCTTTATGAGCCAAATAGAGAT 540
QY 541 CATTGTCGGGCTTGGCCGAGCTTTAGCTTTATCTTAATGCGGCAAAACCTTGGGCGCT 600
DB 541 CATTGTCGGGCTTGGCCGAGCTTTAGCTTTATCTTAATGCGGCAAAACCTTGGGCGCT 600
QY 601 GATGTCATGTAAGCGGCTCTCGAATCTATGCGTTGAAGGGCAATGCTTCTGTAAGCG 660
DB 601 GATGTCATGTAAGCGGCTCTCGAATCTATGCGTTGAAGGGCAATGCTTCTGTAAGCG 660
QY 661 TCGTGTGGGCTGCTTCAACATCCATGATGATGCTTTGATCAGATACGAAAGCAT 720
DB 661 TCGTGTGGGCTGCTTCAACATCCATGATGATGCTTTGATCAGATACGAAAGCAT 720
QY 721 GCGTTGCTTGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 780
DB 721 GCGTTGCTTGGCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 780
QY 781 GTGCGGCTCTTGGCGGAATGAGAGGATTTCTCTACGCAACCTTGATCTGAGTA 840
DB 781 GTGCGGCTCTTGGCGGAATGAGAGGATTTCTCTACGCAACCTTGATCTGAGTA 840
QY 841 CGCATCTCTTGGTAATGAGCGGCAACCTGCTGCTGATTTCCCGTCCGACATTA 900
DB 841 CGCATCTCTTGGTAATGAGCGGCAACCTGCTGCTGATTTCCCGTCCGACATTA 900
QY 901 CGCTTGTCTAATGATGAGCGGCTTAATTAACCGGATGATGAAATGATGATCTTGGT 960
DB 901 CGCTTGTCTAATGATGAGCGGCTTAATTAACCGGATGATGAAATGATGATCTTGGT 960
QY 961 CTTTACGCTTTGGGTAAGCGCTGAGACGCGGTGCGCACTGAGAAATTTGA 1014
DB 961 CTTTACGCTTTGGGTAAGCGCTGAGACGCGGTGCGCACTGAGAAATTTGA 1014
RESULT 3
ADH36219
ID ADH36219 standard; DNA; 1014 BP.
XX
```

```
AC ADH36219;
XX
DT 11-MAR-2004 (first entry)
XX
DE Chemical process monitoring-related nitrilase gene sequence SeqID385.
XX
KM Chemical process monitoring; biochemical process monitoring; cyanide;
XX high throughput system; gene; ds.
XX
OS Unidentified.
XX
PN MO2003098187-A2.
XX
PD 27-NOV-2003.
XX
PF 15-MAY-2003; 2003MO-US015639.
XX
PR 15-MAY-2002; 2002US-0380737P.
XX
PA (DIVE-) DIVERSA CORP.
XX
PI Weiner D, Chaplin JA, Chi E, Milan A, Desantis G, Burk MJ;
PI Mcquaid J, Stege J;
XX
DR WPI; 2004-142708/14.
XX
P-PsDB; ADH36220.
XX
PT Monitoring a chemical or biochemical process comprises providing a
PT reactant comprising a cyanide or a material that can be converted to
PT cyanide or a reactant that generates a cyanide or a material that can be
PT converted to cyanide.
XX
PS Claim 74; SEQ ID NO 385; 277pp; English.
XX
CC This invention relates to a novel method of monitoring chemical or
CC biochemical processes. The method involves providing a reactant
CC comprising cyanide (or a material that can be converted to a cyanide)
CC that generates as a reaction product cyanide or a material that can be
CC converted to cyanide and measuring the concentration of produced cyanide.
CC The method is useful for monitoring a chemical or biochemical process.
CC sensitive to detect a small amount of product. The present sequence is
CC that of a gene which encodes a nitrilase enzyme which can be used in the
CC method of the invention.
XX
SQ Sequence 1014 BP; 238 A; 232 C; 279 G; 265 T; 0 U; 0 Other;
XX
Query Match 99.8%; Score 1012.4; DB 12; Length 1014;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1013; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
QY 1 ATGAAAGAGCTATCAAGGTCGCTGCGTGAAGCCGCCGATCTACATGATTTGAG 60
DB 1 ATGAAAGAGCTATCAAGGTCGCTGCGTGAAGCCGCCGATCTACATGATTTGAG 60
QY 61 GGGACGGTGGACAAACCATTTGATGATGAAAGACACGTAATTAATGCTGCTG 120
DB 61 GGGACGGTGGACAAACCATTTGATGATGAAAGACACGTAATTAATGCTGCTG 120
QY 121 ATGCGCTTTCCGGAACCTTGATTCAGGCTACCCATGGTTTCTTTGGCTGACTGACA 180
DB 121 ATGCGCTTTCCGGAACCTTGATTCAGGCTACCCATGGTTTCTTTGGCTGACTGACA 180
QY 181 GCATGGGCAATGCATTTGTACGCCAATACATGAGACCTCATTTGATGATGAGCCCT 240
DB 181 GCATGGGCAATGCATTTGTACGCCAATACATGAGACCTCATTTGATGATGAGCCCT 240
QY 241 CAAGCTAAGCGCATTTCAATGACGCAAGCGGTTGGAAATCATGGTCACTCTGGGATG 300
DB 241 CAAGCTAAGCGCATTTCAATGACGCAAGCGGTTGGAAATCATGGTCACTCTGGGATG 300
QY 301 AGTGAACGGGTGGTGGCAACCTTTACATCAGTCAATGATGATAGCGGATTAATGCTGAC 360
DB 301 AGTGAACGGGTGGTGGCAACCTTTACATCAGTCAATGATGATAGCGGATTAATGCTGAC 360
```

```
Db 301 AGTGACGGGTGGGTGGACCCCTTACATCATGATGTTTCATAGGCGATAATGATGAC 360
Qy 361 ACCATTGGGGCCCGCGCAAAAGTTGAAACCTACTTTGTTGAACGTACTTTGTTGGCGAA 420
Db 361 ACCATTGGGGCCCGCGCAAAAGTTGAAACCTACTTTGTTGAACGTACTTTGTTGGCGAA 420
Qy 421 GGGGATGGTTCATGCTAGCGGGTTTGGAGACGTCGTGTGGAAAGCGTGGTGGCTTATGC 480
Db 421 GGGGATGGTTCATGCTAGCGGGTTTGGAGACGTCGTGTGGAAAGCGTGGTGGCTTATGC 480
Qy 481 TGTGGGAGACCTTCAACCGCTAACAATAAGCTTTGTATGACAAATGAGAGATT 540
Db 481 TGTGGGAGACCTTCAACCGCTAACAATAAGCTTTGTATGACAAATGAGAGATT 540
Qy 541 CATGTGCGGCTTGGCGGAGCTTATAGCTTTATCTTAATGCGGAAAGCCCTGGGCT 600
Db 541 CATGTGCGGCTTGGCGGAGCTTATAGCTTTATCTTAATGCGGAAAGCCCTGGGCT 600
Qy 601 GATGTCAATGTAAGGGGCTCTCGAATCTATGCCGTTGAAGGGCAATGCTGTACTAGG 660
Db 601 GATGTCAATGTAAGGGGCTCTCGAATCTATGCCGTTGAAGGGCAATGCTGTACTAGG 660
Qy 661 TCGTGTGCGCTCGTTTCAATCCATCATGATCGATATGCTTTGTAGAGATGAGAAAGCAT 720
Db 661 TCGTGTGCGCTCGTTTCAATCCATCATGATCGATATGCTTTGTAGAGATGAGAAAGCAT 720
Qy 721 GCGTGTGCTTGTGCTGTGTGTGACACTCAGCATCATATAGGGCTGATGTGTGACTTG 780
Db 721 GCGTGTGCTTGTGCTGTGTGTGACACTCAGCATCATATAGGGCTGATGTGTGACTTG 780
Qy 781 GTGGGCGCTTGTGCGAAATGAAAGAGGGTATTCCTTAACGCAAACTTGATCTGGAGTA 840
Db 781 GTGGGCGCTTGTGCGAAATGAAAGAGGGTATTCCTTAACGCAAACTTGATCTGGAGTA 840
Qy 841 CGCATCTTGTCTAAATGCGCGAGACCCCTGCTGTCTATATTCCTGCTCCGACATTA 900
Db 841 CGCATCTTGTCTAAATGCGCGAGACCCCTGCTGTCTATATTCCTGCTCCGACATTA 900
Qy 901 CGCTTGTCTAATGATCGGAGCCCTTAATTAACCGGTAGTTGAAATGAAAGGTGATCTT 960
Db 901 CGCTTGTCTAATGATCGGAGCCCTTAATTAACCGGTAGTTGAAATGAAAGGTGATCTT 960
Qy 961 CCTTACGCTTGGGTAAGCGCTGAGAGCGGGTGGCAACTCGAAGAAATTTGA 1014
Db 961 CCTTACGCTTGGGTAAGCGCTGAGAGCGGGTGGCAACTCGAAGAAATTTGA 1014

RESULT 4
ADG93918
ID ADG93918 standard; DNA; 1014 BP.
AC ADG93918;
XX
XX 11-MAR-2004 (first entry)
DT
XX Nitrilase enzyme gene sequence Segid385.
DE
XX Nitrilase; nitrile; carboxylic acid; chemical process; pH; temperature;
KM enantioselective transformation; gene; db.
XX
OS Unidentified.
XX
XX MO2003097810-A2.
XX
XX 27-NOV-2003.
XX
XX 15-MAY-2003; 2003WO-US015712.
XX
XX 15-MAY-2002; 2002US-00146772.
XX
XX 09-SEP-2002; 2002US-00241742.
XX
XX (DIVE-) DIVERSA CORP.
XX
```

```
PI Desantis G, Short JM, Burk M, Wong K, Farwell R, Chatman K;
XX WPI; 2004-090637/09.
DR P-PSDB; ADG93918.
XX
XX New isolated or recombinant nucleic acid encoding a polypeptide having
PT nitrilase activity, useful for screening enantioselective transformation.
XX
XX Claim 44; SEQ ID NO 385; 295pp; English.
XX
XX This invention is related to a novel isolated or recombinant nucleic acid
CC encoding a protein having nitrilase activity. Nitrilase's are capable of
CC converting nitrile's directly to carboxylic acids and have great
CC potential for use in industrial chemical processes. The isolated
CC nitrilase proteins of the invention have increased activity and stability
CC at increased pH and temperature when compared to those conventionally
CC used. In addition, the nucleic acid of the invention is useful for
CC screening enantioselective transformation. The present sequence is that
CC of a DNA sequence which encodes a nitrilase enzyme of the invention.
XX
XX Sequence 1014 BP; 238 A; 232 C; 279 G; 265 T; 0 U; 0 Other;
SQ
Query Match 99.8%; Score 1012.4; DB 12; Length 1014;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1013; Conservative 0; Mismatches 1; Indels 0; Gaps 0;
Qy 1 ATGAAAGAGTATCAAGTTCGCTGGTGAACCCCGGACTCTCATGATTTGGAG 60
Db 1 ATGAAAGAGTATCAAGTTCGCTGGTGAACCCCGGACTCTCATGATTTGGAG 60
Qy 61 GCGACGGTGGCAAAACCATTTGATGATGGAAGAGCAGCCTAATATGCTGCTG 120
Db 61 GCGACGGTGGCAAAACCATTTGATGATGGAAGAGCAGCCTAATATGCTGCTG 120
Qy 121 ATGCGCTTTCGGAACCTGGATTCAGGCTACCCATGCTTTCTTGGCTTGA 180
Db 121 ATGCGCTTTCGGAACCTGGATTCAGGCTACCCATGCTTTCTTGGCTTGA 180
Qy 181 GCATGGGGAATGCAATTTGTATGCGCATATCCATGAGAACTGATTTGGAGTGG 240
Db 181 GCATGGGGAATGCAATTTGTATGCGCATATCCATGAGAACTGATTTGGAGTGG 240
Qy 241 CAAGCTAAGCCGATTTGATGATGAGCAGCAGCGGTTGGAACTCATGTGACCT 300
Db 241 CAAGCTAAGCCGATTTGATGATGAGCAGCAGCGGTTGGAACTCATGTGACCT 300
Qy 301 AGTGAACGGGTGCGTGGCACTTTTACATCATGATGATGATGATGATGATGATG 360
Db 301 AGTGAACGGGTGCGTGGCACTTTTACATCATGATGATGATGATGATGATGATG 360
Qy 361 ACCATTGGGGCCCGCGCAAAAGTTGAAACCTAATTTGTTGAACGTAATTTGTT 420
Db 361 ACCATTGGGGCCCGCGCAAAAGTTGAAACCTAATTTGTTGAACGTAATTTGTT 420
Qy 421 GGGGATGGTTCATGCTAGCGGGTTTGGAGACGTCGTGTGGAAAGCGTGGTGG 480
Db 421 GGGGATGGTTCATGCTAGCGGGTTTGGAGACGTCGTGTGGAAAGCGTGGTGG 480
Qy 481 TGTGGGAGACCTTCAACCGCTAACAATAAGCTTTGTATGACAAATGAGAGATT 540
Db 481 TGTGGGAGACCTTCAACCGCTAACAATAAGCTTTGTATGACAAATGAGAGATT 540
Qy 541 CATGTGCGGCTTGGCGGAGCTTATAGCTTTATCTTAATGCGGAAAGCCCTGG 600
Db 541 CATGTGCGGCTTGGCGGAGCTTATAGCTTTATCTTAATGCGGAAAGCCCTGG 600
Qy 601 GATGTCAATGTAAGGGGCTCTCGAATCTATGCCGTTGAAGGGCAATGCTGT 660
Db 601 GATGTCAATGTAAGGGGCTCTCGAATCTATGCCGTTGAAGGGCAATGCTGT 660
Qy 661 TCGTGTGCGCTCGTTTCAATCCATCATGATCGATATGCTTTGTAGAGATGAG 720
Db 661 TCGTGTGCGCTCGTTTCAATCCATCATGATCGATATGCTTTGTAGAGATGAG 720
```

QY 721 GCGTTGCTTCTGGCTGTGTGGACACTCAGCATATGAGGCGTATGGTGTGACTTG 780  
|||  
DB 721 GCGTTGCTTCTGGCTGTGTGGACACTCAGCATATGAGGCGTATGGTGTGACTTG 780  
|||  
QY 781 GTCGGCGCTCTTGGCCGAAATGAAAGAGGATTTCTTACGCAACCTTGATCCTGAGTA 840  
|||  
DB 781 GTCGGCGCTCTTGGCCGAAATGAAAGAGGATTTCTTACGCAACCTTGATCCTGAGTA 840  
|||  
QY 841 CGCATCTTGTCTAAATATGCGCGAGACCTGCTGTGATATTCCCGTCCGACATTACT 900  
|||  
DB 841 CGCATCTTGTCTAAATATGCGCGAGACCTGCTGTGATATTCCCGTCCGACATTACT 900  
|||  
QY 901 CGCTTGCTAATATGATGCGACCCCTTAATTAACGGGTATGAAATGAGTATCTTCTG 960  
|||  
DB 901 CGCTTGCTAATATGATGCGACCCCTTAATTAACGGGTATGAAATGAGTATCTTCTG 960  
|||  
QY 961 CCTTACGCTTTGGGTAAAGCGTCTGAGACGGGTGCGCAACTCGAAGAAATTTGA 1014  
|||  
DB 961 CCTTACGCTTTGGGTAAAGCGTCTGAGACGGGTGCGCAACTCGAAGAAATTTGA 1014  
|||  
RESULT 5  
AD162517  
ID AD162517 standard; DNA; 1014 BP.  
XX  
AC AD162517;  
XX  
DT 22-APR-2004 (first entry)  
XX  
DE DNA encoding nitrilase polypeptide #193.  
XX  
KM Atorvastatin; (R)-ethyl 4-cyano-3-hydroxybutyrate;  
KM (R)-ethyl 4-cyano-3-hydroxybutyric acid; epichlorohydrin;  
KM 3-hydroxyglutaronitrile; 3-hydroxyglutaronitrile;  
KM 4-cyano-3-hydroxybutyric acid; ethyl-4-cyano-3-hydroxybutyric acid;  
KM mixed hyperlipidaemia; homozygous familial hypercholesterolaemia;  
KM antilipaeamic; gene; ds.  
XX  
OS Unidentified.  
XX  
PN WO2003106415-A2.  
XX  
PD 24-DEC-2003.  
XX  
PF 13-JUN-2003; 2003WO-US018840.  
XX  
PR 13-JUN-2002; 2002US-0389317P.  
XX  
PR 28-JUN-2002; 2002US-0392944P.  
XX  
PA (DIVE-) DIVERSA CORP.  
XX  
PI Burk M, Desantis G, Morgan B, Zhu Z;  
PI MPI: 2004-090821/09.  
DR P-PSDB; AD162518.  
XX  
PT Preparation of atorvastatin comprises catalytic conversion of 3-  
PT hydroxyglutaronitrile by polypeptide with nitrilase activity, converting  
PT obtained 4-cyano-3-hydroxybutyric acid to ethyl-4-cyano-3-hydroxybutyric  
PT acid and forming atorvastatin.  
XX  
PS Claim 46; SEQ ID NO 385; 253pp; English.  
XX  
CC The present invention relates to a method for preparing an atorvastatin  
CC intermediate known as (R)-ethyl 4-cyano-3-hydroxybutyrate ((R)-ethyl 4-  
CC cyano-3-hydroxybutyric acid). The method comprises optionally converting  
CC epichlorohydrin or equivalent to 3-hydroxyglutaronitrile, catalytic  
CC conversion of 3-hydroxyglutaronitrile or equivalent to 4-cyano-3-  
CC hydroxybutyric acid with a polypeptide having nitrilase activity,  
CC converting 4-cyano-3-hydroxybutyric acid to ethyl-4-cyano-3-hydroxybutyric  
CC acid, and converting this to (R)-ethyl 4-cyano-3-hydroxybutyrate. The  
CC method involves whole cell processes, cell lysate process, "one pot"

CC processes, and "multi-pot" processes using a variety of parameters.  
CC Atorvastatin is used, in conjunction with dietary restriction, in the  
CC management of hyperlipidaemia, including hypercholesterolemia, mixed  
CC dyslipidaemia and homozygous familial hypercholesterolemia. The present  
CC sequence encodes a nitrilase polypeptide obtained from an environmental  
CC sample.  
SQ Sequence 1014 BP; 238 A; 232 C; 279 G; 265 T; 0 U; 0 Other;  
Query Match 99.8%; Score 1012.4; DB 12; Length 1014;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 1013; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 ATGAAAGAGCTATCAAGTCCCGCTGCAAGCCGCCGATCTACATGATTTGGAG 60  
|||  
DB 1 ATGAAAGAGCTATCAAGTCCCGCTGCAAGCCGCCGATCTACATGATTTGGAG 60  
|||  
QY 61 GCGACGGTGAACAAACCAATTGATGAGAAAGACAGCAGTATATGCTCGTCTG 120  
|||  
DB 61 GCGACGGTGAACAAACCAATTGATGAGAAAGACAGCAGTATATGCTCGTCTG 120  
|||  
QY 121 ATCGCCTTTCCGAAAATTGATTCAGGCTACCAATGCTTTCTTGGCTTGACTACCA 180  
|||  
DB 121 ATCGCCTTTCCGAAAATTGATTCAGGCTACCAATGCTTTCTTGGCTTGACTACCA 180  
|||  
QY 181 GCATGGGCAATGCAATTTGTATGCCCAATACATGAGAACTCATTTGAGTTGATGGCCCT 240  
|||  
DB 181 GCATGGGCAATGCAATTTGTATGCCCAATACATGAGAACTCATTTGAGTTGATGGCCCT 240  
|||  
QY 241 CAAGCTAAGCGCATTTTCAGATGACGACGAGGGGTGGGAATAGTGTACCCGCGGGATG 300  
|||  
DB 241 CAAGCTAAGCGCATTTTCAGATGACGAGGGGTGGGAATAGTGTACCCGCGGGATG 300  
|||  
QY 301 AGTGAACGGGTGTGTGCAACCTTTATCATCATGATGATGATGATGATGATGATGATG 360  
|||  
DB 301 AGTGAACGGGTGTGTGCAACCTTTATCATCATGATGATGATGATGATGATGATGATG 360  
|||  
QY 361 ACCATTGGGGGCGCGCGAAGTTGAACCTACTTTTGTGAACGTAATCTTTGGCGGAA 420  
|||  
DB 361 ACCATTGGGGGCGCGCGAAGTTGAACCTACTTTTGTGAACGTAATCTTTGGCGGAA 420  
|||  
QY 421 GGGGATGGTTCATCGCTAGCGGTTTTCGAGACGTCGTTGGAAGGTGGGTGCTTATGC 480  
|||  
DB 421 GGGGATGGTTCATCGCTAGCGGTTTTCGAGACGTCGTTGGAAGGTGGGTGCTTATGC 480  
|||  
QY 481 TGTGGAGACACTTTCACACCGCTAACAAATACGCTTTGTATGCAAAATGAAAGATT 540  
|||  
DB 481 TGTGGAGACACTTTCACACCGCTAACAAATACGCTTTGTATGCAAAATGAAAGATT 540  
|||  
QY 541 CATTTGTGGGCTTGGCCGAGCTTTAGCCTTATCTTAATGCGGCGAAGCCCTGGGGCT 600  
|||  
DB 541 CATTTGTGGGCTTGGCCGAGCTTTAGCCTTATCTTAATGCGGCGAAGCCCTGGGGCT 600  
|||  
QY 601 GATGTCATATGATGCGGCTTCGATCTATGCGGTGAAGGCGAATGCTGCTACTAGGG 660  
|||  
DB 601 GATGTCATATGATGCGGCTTCGATCTATGCGGTGAAGGCGAATGCTGCTACTAGGG 660  
|||  
QY 661 TGTGTGGCTGCTGTTTCAATTCATGATGATGATGATGATGATGATGATGATGATGAT 720  
|||  
DB 661 TGTGTGGCTGCTGTTTCAATTCATGATGATGATGATGATGATGATGATGATGATGAT 720  
|||  
QY 721 GCGTTGCTTCTGGCTGTGTGGACACTCAGCATATGAGGCGTATGGTGTGACTTG 780  
|||  
DB 721 GCGTTGCTTCTGGCTGTGTGGACACTCAGCATATGAGGCGTATGGTGTGACTTG 780  
|||  
QY 781 GTCGGCGCTCTTGGCCGAAATGAAAGAGGATTTCTTACGCAACCTTGATCCTGAGTA 840  
|||  
DB 781 GTCGGCGCTCTTGGCCGAAATGAAAGAGGATTTCTTACGCAACCTTGATCCTGAGTA 840  
|||  
QY 841 CGCATCTTGTCTAAATATGCGCGAGACCTGCTGTGATATTCCCGTCCGACATTACT 900  
|||  
DB 841 CGCATCTTGTCTAAATATGCGCGAGACCTGCTGTGATATTCCCGTCCGACATTACT 900  
|||

QY 901 CGCTTGCTAATAGATCGACGCTTAAATTACCGGTAGTGAATTAAGTGATCTTCGT 960  
DB 901 CGCTTGCTAATAGATCGACGCTTAAATTACCGGTAGTGAATTAAGTGATCTTCGT 960  
QY 961 CCTTAAGCTTTGGGTAAAGCCTCGAGACGGGTGCGCAACCTCGAAGAAATTGA 1014  
DB 961 CCTTAAGCTTTGGGTAAAGCCTCGAGACGGGTGCGCAACCTCGAAGAAATTGA 1014  
RESULT 6  
ADI64638  
ID ADI64638 standard; DNA; 1014 BP.  
AC ADI64638;  
XX 22-APR-2004 (first entry)  
XX DNA encoding nitrilase seq id 193.  
XX (R)-ethyl 4-cyano-3-hydroxybutyric acid; nitrile hydrolysis;  
KM carboxylic acid; cyanohydrin moiety hydrolysis;  
KM aminonitrile moiety hydrolysis; chiral alpha-hydroxy acid molecule;  
KM chiral amino acid molecule; (S)-ethyl 4-cyano-3-hydroxybutyric acid;  
KM (R)-mandelic acid; (S)-mandelic acid; (S)-phenyl lactic acid derivative;  
KM (R)-phenyl lactic acid derivative; & enantiomeric excess;  
KM & diastomeric excess; food additive; drug intermediate; de; nitrilase;  
gene.  
XX Unidentified.  
XX OS US2004014195-A1.  
XX PN 22-JAN-2004.  
XX PD 15-MAY-2003; 2003US-00440523.  
XX PF 29-DEC-1999; 99US-0173609P.  
XX PR 07-DEC-2000; 2000US-0254414P.  
XX PR 28-DEC-2000; 2000US-00751289.  
XX PR 21-JUN-2001; 2001US-0300189P.  
XX PR 30-JUL-2001; 2001US-030906P.  
XX PR 22-JAN-2002; 2002US-0351336P.  
XX PR 15-MAY-2002; 2002US-00146772.  
XX PR 09-SEP-2002; 2002US-00241742.  
XX (DIVE-) DIVERSA CORP.  
XX PA Desantis G, Short JM, Burk MJ, Wong K, Farwell R, Chatman K;  
XX WIPI: 2004-121569/12.  
XX DR P-PSDB; ADI64639.  
XX PT Novel isolated or recombinant polypeptide having nitrilase activity,  
XX useful in production of food additives.  
XX PS Claim 1; SEQ ID NO 385; 105bp; English.  
XX The invention describes an isolated or recombinant polypeptide (I)  
XX comprising amino acids having a sequence at least 50 % identical to a  
XX sequence (SI) available in electronic form (EC) from the following web  
XX site ftp.segdata.uspto.gov/sequence.html?DocID=2004014195, or its  
XX variants, having one or more mutations at residue 55 Lys, Gly or Glu, at  
XX residue 60 glutamic acid, at residue 111 Ser, their combinations or  
XX fragments. (I) is useful for: producing an (R)-ethyl 4-cyano-3-  
XX hydroxybutyric acid; hydrolysing a nitrile to a carboxylic acid;  
XX hydrolysing cyanohydrin moiety or an aminonitrile moiety; producing a  
XX chiral alpha-hydroxy acid molecule or a chiral amino acid molecule;  
XX producing an (S)-ethyl 4-cyano-3-hydroxybutyric acid; producing an (R)-  
XX mandelic acid or (S)-mandelic acid; producing (S)-phenyl lactic acid  
XX derivative or an (R)-phenyl lactic acid derivative; modifying a molecule;  
XX and for identifying a modified compound. The inventive method is useful  
XX for monitoring or determining & enantiomeric excess or & diastomeric  
XX excess. (I) is useful in the production of food additives and drug

CC intermediates. This sequence encodes a nitrilase of the invention.  
XX  
SQ Sequence 1014 BP; 238 A; 232 C; 279 G; 265 T; 0 U; 0 Other;  
Query Match 99.8%; Score 1012.4; DB 12; Length 1014;  
Best Local Similarity 99.9%; Pred. No. 0;  
Matches 1013; Conservative 0; Mismatches 1; Indels 0; Gaps 0;  
QY 1 ATGAAAGAGCTATCAAGGTGCTGCTGCGTGCAGCCGCCCGATCTACATGATTTGGAG 60  
DB 1 ATGAAAGAGCTATCAAGGTGCTGCTGCGTGCAGCCGCCCGATCTACATGATTTGAAG 60  
QY 61 GCGACGGTGCACAAACCAATGAGTGAATGGAAGAGCAGCATTAATATCTGCTG 120  
DB 61 GCGACGGTGCACAAACCAATGAGTGAATGGAAGAGCAGCATTAATATCTGCTG 120  
QY 121 ATGCGCTTTCCGGAACCTTGATTCAGGCTACCCATGTTCTTTGGCTTGACTCACA 180  
DB 121 ATGCGCTTTCCGGAACCTTGATTCAGGCTACCCATGTTCTTTGGCTTGACTCACA 180  
QY 181 GCATGGGCAATGCAATTTGTAAGCCAAATACCATAGAACTCATTTGAGTTGGATGCCCT 240  
DB 181 GCATGGGCAATGCAATTTGTAAGCCAAATACCATAGAACTCATTTGAGTTGGATGCCCT 240  
QY 241 CAAGCTAAGGSCATTTGAGATGAGCCAGGCGTTGGGAATCATGTCACCCCTGGGAGT 300  
DB 241 CAAGCTAAGGSCATTTGAGATGAGCCAGGCGTTGGGAATCATGTCACCCCTGGGAGT 300  
QY 301 AGTGAAAGGCTCGGTGCGACCCCTTTACATCATGTCAGTGGTTCATAGGCAATATGTTGAC 360  
DB 301 AGTGAAAGGCTCGGTGCGACCCCTTTACATCATGTCAGTGGTTCATAGGCAATATGTTGAC 360  
QY 361 ACCATTGGGGCCCGGCGGAAATGTAACCTACTTTGTTGAACGTACTTTTGGGGGAA 420  
DB 361 ACCATTGGGGCCCGGCGGAAATGTAACCTACTTTGTTGAACGTACTTTTGGGGGAA 420  
QY 421 GGGGATGGTTATGCTAGCTAGCGGTTTGCAGACGCTCTGTGGAAGGCTGGGCTTATGTC 480  
DB 421 GGGGATGGTTATGCTAGCTAGCGGTTTGCAGACGCTCTGTGGAAGGCTGGGCTTATGTC 480  
QY 481 TGTGGAGACACCTTCAACCGCTTAACAAATAGCCTTTGTAATGCAAAATGAAGAGATT 540  
DB 481 TGTGGAGACACCTTCAACCGCTTAACAAATAGCCTTTGTAATGCAAAATGAAGAGATT 540  
QY 541 CATGTGGGCTTGGCCGAGGCTTATGCTTATATCTTAATGCGGGAAGCCCTGGGGCT 600  
DB 541 CATGTGGGCTTGGCCGAGGCTTATGCTTATATCTTAATGCGGGAAGCCCTGGGGCT 600  
QY 601 GATGTCATGTAGGCGGCTCTCGAATCTATCCGTGAAGGCAATGCTTGTAACG 660  
DB 601 GATGTCATGTAGGCGGCTCTCGAATCTATCCGTGAAGGCAATGCTTGTAACG 660  
QY 661 TGTGTGGCTGCTTTCAATTCATGATGATGCTTTGTACATGATGAGAAAGCAT 720  
DB 661 TGTGTGGCTGCTTTCAATTCATGATGATGCTTTGTACATGATGAGAAAGCAT 720  
QY 721 GCGTTGCTTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 780  
DB 721 GCGTTGCTTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 780  
QY 781 GTGCGGCTCTTGGCGGAAATGAAGAGGATTTCTCTACGCAACCTTGATCTGAGATA 840  
DB 781 GTGCGGCTCTTGGCGGAAATGAAGAGGATTTCTCTACGCAACCTTGATCTGAGATA 840  
QY 841 GCGATCTTGGCTTAAATGGCGGAGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 900  
DB 841 GCGATCTTGGCTTAAATGGCGGAGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 900  
QY 901 CGCTTGCTAATAGATCGACGCTTAAATTACCGGTAGTGAATTAAGTGATCTTCGT 960  
DB 901 CGCTTGCTAATAGATCGACGCTTAAATTACCGGTAGTGAATTAAGTGATCTTCGT 960  
QY 961 CCTTAAGCTTTGGGTAAAGCCTCGAGACGGGTGCGCAACCTCGAAGAAATTGA 1014



Db 961 CCTTACGCTTGGGTAAAGGCTGTGAGACGGGTGCCCACTCGAAGAAATTGA 1014

## RESULT 7

ADCC3780  
ID ADCC3780 standard; DNA; 1014 BP.

AC ADCC3780;

DT 18-DEC-2003 (first entry)

DE DNA sequence (SeqID 47) encoding a nitrilase enzyme.

KW gene; ds; nitrilase; nitrile; cyanohydrin; ammonia; biocatalyst;

enantiomer; chiral medicine.

OS Unidentified.

PN WO200300840-A2.

PD 03-JAN-2003.

PF 15-MAY-2002; 2002WO-US015983.

PR 21-JUN-2001; 2001US-0300189P.

PR 30-JUL-2001; 2001US-0309006P.

PR 22-JAN-2002; 2002US-0351336P.

PA (DIVE-) DIVERSA CORP.

PA (MADD/) MADDEN D.

PI Madden JM, Desantis G, Chaplin JA, Weiner DP, Milan A, Chi E,

PI Short JM, Burk M,

DR WPI; 2003-201417/19.

P-PSDB; ADCC3781.

PS Claim 1; SEQ ID NO 47; 560bp; English.

CC This invention relates to nitrilases and the nucleic acids that encode  
CC these enzymes thereof. Specifically, it refers to polypeptides that  
CC exhibit nitrilase activity, i.e. the ability to directly hydrolyse  
CC nitriles or cyanohydrins into their corresponding carboxylic acids and  
CC ammonia. Nitrilases have commercial utility as biocatalysts for use in  
CC the synthesis of enantiomerically pure aromatic and aliphatic amino  
CC acids, as well as hydroxy acids, which are important for the development  
CC of chiral medicines. Furthermore, the present invention describes  
CC nitrilases, isolated from mesophilic microorganisms, that have improved  
CC activity and stability at increased pH and temperature. They are also  
CC are capable of chiral differentiation. This polynucleotide is a DNA  
CC sequence that encodes a nitrilase enzyme of the invention.

XX Sequence 1014 BP; 230 A; 304 C; 284 G; 196 T; 0 U; 0 Other;

Query Match 49.3%; Score 499.4; DB 10; Length 1014;

Best Local Similarity 68.3%; Pred. No. 11e-152;

Matches 692; Conservative 0; Mismatches 321; Indels 0; Gaps 0;

QY 2 TGAAGAAGCTATCAAGCTGCTGCAAGCCGCCATCTACATGATTTGAGG 61

Db 2 TGAAGAAGCTATCAAGCTGCTGCAAGCCGCCATCTACATGATTTGAGG 61

QY 62 CGACGCTGACAAACCTTGAAGTTGATGAAAGACGACGTAATATGCTGCTGA 121

Db 62 CGACGCTGACAAACCTTGAAGTTGATGAAAGACGACGTAATATGCTGCTGA 121

QY 122 TCGCCTTTCCGGAACCTTGATTCAGAGCTACCAATGTTTCTTTGACTGACCGAG 181

Db 122 TCGCCTTTCCGGAACCTTGATTCAGAGCTACCAATGTTTCTTTGACTGACCGAG 181

QY 182 CATGGCAATGCAATTTGTACCGCAATACATAGAACTCATTTGAGTTGATGGCCCTC 241

Db 182 CATGGCAATGCAATTTGTACCGCAATACATAGAACTCATTTGAGTTGATGGCCCTC 241

QY 242 AAGCTAAGCGCATTTTCAGATGCAAGCCGCTTGGAAATCATGGTCACTCTGGAGTGA 301

Db 242 AAGCTAAGCGCATTTTCAGATGCAAGCCGCTTGGAAATCATGGTCACTCTGGAGTGA 301

QY 302 GTGAACGGGTGGGTGGACCCCTTTCATCAGTCAGTGGTTCATAGCGCATATGAGTGA 361

Db 302 GTGAACGGGTGGGTGGACCCCTTTCATCAGTCAGTGGTTCATAGCGCATATGAGTGA 361

QY 302 GCGAAGCGGTGAGCGGAACCTTCTACATGGGCAAGTGGCTCATTAACGATTAAGGCGA 361

Db 302 GCGAAGCGGTGAGCGGAACCTTCTACATGGGCAAGTGGCTCATTAACGATTAAGGCGA 361

QY 362 CCATTGGGCGCCCGGCGAAGTTGAAACCTACTTTTGTGAAAGTACTTTTGTGGCGAAG 421

Db 362 CCATTGGGCGCCCGGCGAAGTTGAAACCTACTTTTGTGAAAGTACTTTTGTGGCGAAG 421

QY 422 GCGATGCTTCATCGCTAGCGGTTTTCAGACGCTGTTGAAAGCTGGGCTTATGCT 481

Db 422 GCGATGCTTCATCGCTAGCGGTTTTCAGACGCTGTTGAAAGCTGGGCTTATGCT 481

QY 482 GTTGGAGACACCTTCAACCGCTAACAAATAGCTTTGTATGACAAATAGAGATTC 541

Db 482 GTTGGAGACACCTTCAACCGCTAACAAATAGCTTTGTATGACAAATAGAGATTC 541

QY 542 ATTGTCGCGCTTGGCGGAGCTTTAGCCTTTATCTTAATGCGCGAAGCCCTGGGCGCTG 601

Db 542 ATTGTCGCGCTTGGCGGAGCTTTAGCCTTTATCTTAATGCGCGAAGCCCTGGGCGCTG 601

QY 542 ACTTGGCGCGCTGGCTGATGCTTACGACATCTACCGTCAAGCAGACGATCTTGAACGAG 601

Db 542 ACTTGGCGCGCTGGCTGATGCTTACGACATCTACCGTCAAGCAGACGATCTTGAACGAG 601

QY 602 ATGTCAATGATGAGCGGCTCTGTGAATCTATGCGCTTGAAGGCGAATGCTTGTACTAGCGT 661

Db 602 ATGTCAATGATGAGCGGCTCTGTGAATCTATGCGCTTGAAGGCGAATGCTTGTACTAGCGT 661

QY 602 AAGTAAATGTGCGAGCTTCTGGAATCTAGCGCGTGAAGGCGAAGTGTGTTCTGCGCTT 661

Db 602 AAGTAAATGTGCGAGCTTCTGGAATCTAGCGCGTGAAGGCGAAGTGTGTTCTGCGCTT 661

QY 662 CGTGTGCGCTGCTTCAACATCCATGATGATGATGATGATGATGATGATGATGATGATG 721

Db 662 CGTGTGCGCTGCTTCAACATCCATGATGATGATGATGATGATGATGATGATGATGATG 721

QY 722 CGTTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 781

Db 722 CGTTGCTTCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 781

QY 782 TCGGCGCTTTCGCGAAGGATGATGATGATGATGATGATGATGATGATGATGATGATG 841

Db 782 TCGGCGCTTTCGCGAAGGATGATGATGATGATGATGATGATGATGATGATGATGATG 841

QY 842 GATTCCTTGTAAATAGCGGCAAGCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 901

Db 842 GATTCCTTGTAAATAGCGGCAAGCCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 901

QY 902 GCTTCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 961

Db 902 GCTTCTTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 961

QY 962 CTTAAGCTTGGGTAAAGCGTCTGAGACGGGTGCGCACTTGAAGAAATTTGA 1014

Db 962 CATCGGCCAAGGCTTTGAAGTGAAGGCGGCCCGGAGTACGAAAGGCGATTTGA 1014

## RESULT 8

ADH35881  
ID ADH35881 standard; DNA; 1014 BP.

AC ADH35881;

DT 11-MAR-2004 (first entry)

DE Chemical process monitoring-related nitrilase gene sequence SeqID47.

Db 11-MAR-2004 (first entry)

KW high throughput system; gene; ds.

XX Unidentified.  
 OS W02003098187-A2.  
 XX  
 XX 27-NOV-2003.  
 PD  
 XX 15-MAY-2003; 2003WO-US015639.  
 PF  
 XX 15-MAY-2002; 2002US-0380737P.  
 PR  
 XX (DIVE-) DIVERSA CORP.  
 PA  
 XX Weiner D, Chaplin JA, Chi E, Milan A, Desantis G, Burk MJ;  
 PI Mcguaid J, Stege J;  
 XX  
 XX WPI; 2004-142708/14.  
 DR P-PSDB; ADH35882.  
 XX  
 XX Monitoring a chemical or biochemical process comprises providing a  
 PT reactant comprising a cyanide or a material that can be converted to  
 PT cyanide or a reactant that generates a cyanide or a material that can be  
 PT converted to cyanide.  
 PS Claim 74; SEQ ID NO 47; 277bp; English.  
 XX  
 XX This invention relates to a novel method of monitoring chemical or  
 CC biochemical processes. The method involves providing a reactant  
 CC comprising cyanide (or a material that can be converted to a cyanide)  
 CC that generates as a reaction product cyanide or a material that can be  
 CC converted to cyanide and measuring the concentration of produced cyanide.  
 CC The method is useful for monitoring a chemical or biochemical process.  
 CC The method is effective for high throughput systems and is sufficiently  
 CC sensitive to detect a small amount of product. The present sequence is  
 CC that of a gene which encodes a nitrilase enzyme which can be used in the  
 CC method of the invention.  
 CC  
 XX Sequence 1014 BP, 230 A; 304 C; 284 G; 196 T; 0 U; 0 Other;  
 SQ  
 Query Match 49.3%; Score 499.4; DB 12; Length 1014;  
 Best Local Similarity 68.3%; Pred. No. 1.1e-152;  
 Matches 692; Conservative 0; Mismatches 321; Indels 0; Gaps 0;

DB 422 GCGACGAGTATCCTTTCACATTTGACAGACCGTTGGGGGCTGTGGCGGACTGTCT 481  
 QY 482 GTTGGAGACACTTGAACCGCTTAACAAATAGCCTTGTATGCAACAAATGAAGAGTTC 541  
 DB 482 GTTGGAGACACTTGAACCGCTTTCGAAATATGCGCTTACGACAGAAAGGAAATAC 541  
 QY 542 ATTGTGGGCTTGGCGGAGCTTATAGCCTTATCTTAATGCGGAAAGCCTGGGGCTG 601  
 DB 542 ACTTGGCGGCTTGGCGGAGCTTATAGCCTTATCTTAATGCGGAAAGCCTGGGGCTG 601  
 QY 602 ATGTCAATGTAGCGGCTCTGAAATCTATGCGCTTGAAGGCAATGCTTGTACTAGCGT 661  
 DB 602 AAGTAAATGTGCGAGCTTCTCGATCTACCGCTGGAAGGCGAGTGTGTCTGCTT 661  
 QY 662 CGTGTGCTCTGCTTTCACATTCATGATGATGATGATGATGATGATGATGATGATGATG 721  
 DB 662 CTTGGCGGCTGCTGCTGCGGAGATGATGATGATGATGATGATGATGATGATGATGATG 721  
 QY 722 CGTGTGCTTGGGCTGATGACACTCACTCATATAGGAGCTGATGATGATGATGATGATG 781  
 DB 722 GCTTCTTCAAGCGCGGCGGCGGATCTCGGCAATATCGGTCGCGATGCGAGCACTAG 781  
 QY 782 TCGGCGCTCTTCCGAAATGAAGGGTATTTCTTACGCAACCTTGAATCTTGAATAC 841  
 DB 782 CGCGCCCTTGGGCGAAAGAGAGATGATTTCTTATGCACTTGAACCTTGGCCGCTC 841  
 QY 842 GCATCCTTGTAAATAGCGGCGAGACCTGTGCTATTAATCCGTCGCACTTACTC 901  
 DB 842 GAATCTTATGCAAGACCGAGCTGATTCAGCCGGGCACTTCAAGACAGACCTACTC 901  
 QY 902 GCTTGTAAATAGATCGAGCCCTAATTAATCCGATGTTGAATGAAGTGAATCTTGTCTC 961  
 DB 902 GCGCTGATCAATCGAGATGCCAATCAGCCAGTCGTAAGAGTGTGAAGGAAATACCTG 961  
 QY 962 CTTAAGCTTGGGTTAAAGCGCTTGAACCGGCTGCAACGGTGGCAACTGGAAGAAATTGA 1014  
 DB 962 CATCGGCCAAGGCTTTGAAGTGAAGCGGCGCCCGGATACGAAGCGAATTGA 1014

RESULT 9  
 ADG93582  
 ID ADG93582 standard; DNA; 1014 BP.  
 XX  
 AC ADG93582;  
 XX  
 DT 11-MAR-2004 (first entry)  
 XX  
 DE Nitrilase enzyme gene sequence SeqId#7.  
 XX  
 KW nitrilase; nitrile; carboxylic acid; chemical process; pH; temperature;  
 KW enantioselective transformation; gene; ds.  
 XX  
 OS Unidentified.  
 XX  
 PN W02003097810-A2.  
 PD 27-NOV-2003.  
 XX  
 PD 15-MAY-2003; 2003WO-US015712.  
 PF  
 XX 15-MAY-2002; 2002US-00146772.  
 PR 09-SEP-2002; 2002US-00241742.  
 XX  
 XX (DIVE-) DIVERSA CORP.  
 PA  
 XX Desantis G, Short JM, Burk M, Mong K, Farwell R, Chatman K;  
 PI  
 XX WPI; 2004-090637/09.  
 DR P-PSDB; ADG93583.  
 XX  
 XX New isolated or recombinant nucleic acid encoding a polypeptide having  
 PT nitrilase activity, useful for screening enantioselective transformation.  
 XX

PS Claim 44; SEQ ID NO 47; 295bp; English.

XX This invention is related to a novel isolated or recombinant nucleic acid  
CC encoding a protein having nitrilase activity. Nitrilase's are capable of  
CC converting nitrile's directly to carboxylic acids and have great  
CC potential for use in industrial chemical processes. The isolated  
CC nitrilase proteins of the invention have increased activity and stability  
CC at increased pH and temperature when compared to those conventionally  
CC used. In addition, the nucleic acid of the invention is useful for  
CC screening enantioselective transformation. The present sequence is that  
CC of a DNA sequence which encodes a nitrilase enzyme of the invention.

XX Sequence 1014 BP; 230 A; 304 C; 284 G; 196 T; 0 U; 0 Other;

Query Match 49.3%; Score 499.4; DB 12; Length 1014;

Best Local Similarity 68.3%; Pred. No. 1.1e-152;

Matches 692; Conservative 0; Mismatches 321; Indels 0; Gaps 0;

QY 2 TGAAGAGCATATCAAGGTCGCTGCGCAAGCCGCCGATCTACATGATTTGGAGG 61  
DB 2 TGAAGAGCATATCAAGGTCGCTGCGCAAGCCGCCGATCTACATGATTTGGAGG 61  
QY 62 CGACGGTGGACAAACCATTTGATGATGAGAGACGACGTAATTAATGCTGCTGA 121  
DB 62 CCACAGTGGACAAACCATTTGATGAGAGACGACGCTTAACGCGCACGCTAA 121  
QY 122 TCGGCTTCCGGAACCTTGAATTCAGGCTCCCATGGTTCTTTGGCTTGAATCACCAG 181  
DB 122 TCGGCTTCCGGAACCTTGAATTCAGGCTCCCATGGTTCTTTGGCTTGAATCACCAG 181  
QY 182 CATGGGCAATGCAATTTGATGCGCAATATCATGAGAACTCATTTGAGTGGAGCCCTC 241  
DB 182 CTGGGGGATGCAATTTGATGCGCGATATCACGAGAACTCATGCTCTGACAGCCCTC 241  
QY 242 AAGCTAAGCGCATTTCAAGTCAAGCCCAAGCGGTTGGAAATCATGTCACCTTGGGATGA 301  
DB 242 AAGCGCAAGCGCATATGATGAGGCGCCGCGCGGTAATATGCTGCGCTGAGGATGA 301  
QY 302 GTGAACGGGTCGGTGGGACCTTTATCATCATGTCATGCTCATAGGGAATATGATGACA 361  
DB 302 GCGAAGCGGTCGGTGGGACCTTTATCATCATGTCATGCTCATAGGGAATATGAGGCAAA 361  
QY 362 CCATTTGGGGCCCGCGCAAGTTGAAACCTACTTTTGTGAACTGTTGTTGGCGGCAAG 421  
DB 362 CAGCTGGGCTGCGCGCAAGTGAACCAACCATGATGAGCGAACCTCTTCGCTGAAG 421  
QY 422 GGAATGTTATCGCTAAGCGGTTTTCAGACGTCGTGTTGAAGGCTGGGCTTATGCT 481  
DB 422 GCGACGATCATCCCTTTCACCTTTCAGACACCGTTGGGGGTGCTGGGCGGACTGCT 481  
QY 482 GTTGGAGCACTTTCACCGCTAACAAATACGTTTGTATGCAAAATGAAAGATTC 541  
DB 482 GTTGGAGCACTTTCACCGCTTTCGAAATATGAGGCTTACGCAAGAAAGAAATAC 541  
QY 542 ATTGTGGGGCTTGGCCGAGCTTTAGCCTTATCCTAATGCGGGAAGCCCTGGGGGCTG 601  
DB 542 ACTTGGCGGCTGGCTTACGCTTACGATCTACGCTCAAGGACAGAAAGTCTTTGACACG 601  
QY 602 ATGTCAATGATGAGGCGCTCTGATCTATGCGGTTGAAGGCAATGCTTGTACTAGCT 661  
DB 602 AAGTAATGTCGAGCTTCTCGATCTAAGCCGCGGAAAGGAGTGTGTTGTTCTGCTT 661  
QY 662 CGTGTGGCTTCGCTTTCACATCATGATGATGCTTTTGTACAGATGAGAAAGATG 721  
DB 662 CCGTGGGCTTCGCTTTCGCGAGATGATGAAATGCTCTCAGCTGAGAAAGCAAGACA 721  
QY 722 CGTTGCTTTCGCTGCTGCTGAGCACTCAAGTATCATAGGCGCTGATGATGCTTGG 781  
DB 722 GCTTCTTTCAGGCGGCGGCGGAGTCTCCGCAATTCGATCCGATGGCAGCACTTAG 781  
QY 782 TCGCGCTCTTTCGCGAAATGAAGAGGTAATCTCTAAGCAAACTTGATCTGAGTAC 841  
DB 782 CGGCGCTTTCGCGGAAAGCAAGAGGTAATCTCTAAGCAAACTTGATCTGAGTAC 841

QY 842 GATCCTTGTCTAAATGAGCGGACAGACCTGCTGCTCATTTATTCGCGTCCGACATTA 901  
DB 842 GATCCTTGTCTAAATGAGCGGACAGACCTGCTGCTCATTTATTCGCGTCCGACATTA 901  
QY 902 GCTTGTCTAATGATCGAGCCCTTAATTAACGCTGATGTTGAATTTGAAGTATCTTCTC 961  
DB 902 GCTTGTCTAATGATCGAGCCCTTAATTAACGCTGATGTTGAATTTGAAGTATCTTCTC 961  
QY 962 CTTACGCTTTGGGTTAAAGCGCTTGAACCGGCTGCGCAACTGGAAGAAATTTGA 1014  
DB 962 CATGGCCCAAGGCTTTGAAGTTGAGCGGCGCCCGGATGCAAGAGCGCATTTGA 1014

#### RESULT 10

AD162179  
ID AD162179 standard; DNA, 1014 BP.

AC AD162179;

DT 22-APR-2004 (first entry)

XX DNA encoding nitrilase polypeptide #24.

DE Atorvastatin; (R)-ethyl 4-cyano-3-hydroxybutyrate;

XX (R)-ethyl 4-cyano-3-hydroxybutyric acid; epichlorohydrin;

KW 3-hydroxyglutaronitrile; 3-hydroxyglutaronitrile;

KW 4-cyano-3-hydroxybutyric acid; ethyl-4-cyano-3-hydroxybutyric acid;

KW mixed hyperlipidaemia; homozygous familial hypercholesterolaemia;

KW antilipemic; gene; ds.

XX Unidentified.

XX WO2003106415-A2.

XX 24-DEC-2003.

PF 13-JUN-2003; 2003MO-US018840.

XX 13-JUN-2002; 2002US-0389317P.

PR 28-JUN-2002; 2002US-0392944P.

XX (DIVE-) DIVERSA CORP.

PA Burk M, Desantis G, Morgan B, Zhu Z;

PI WPI; 2004-090821/09.

DR P-PSDB; AD162180.

DR Preparation of atorvastatin comprises catalytic conversion of 3-

PT hydroxyglutaronitrile by polypeptide with nitrilase activity, converting

PT obtained 4-cyano-3-hydroxybutyric acid to ethyl-4-cyano-3-hydroxybutyric

PT acid and forming atorvastatin.

XX Claim 46; SEQ ID NO 47; 253bp; English.

PS The present invention relates to a method for preparing an atorvastatin

XX intermediate known as (R)-ethyl 4-cyano-3-hydroxybutyrate ((R)-ethyl 4-

CC cyano-3-hydroxybutyric acid). The method comprises optionally converting

CC epichlorohydrin or equivalent to 3-hydroxyglutaronitrile, catalytic

CC conversion of 3-hydroxyglutaronitrile or equivalent to 4-cyano-3-

CC hydroxybutyric acid with a polypeptide having nitrilase activity,

CC converting 4-cyano-3-hydroxybutyric to ethyl-4-cyano-3-hydroxybutyric

CC acid, and converting this to (R)-ethyl 4-cyano-3-hydroxybutyrate. The

CC method involves whole cell processes, cell lysate processes, "one pot"

CC processes, and "multi-pot" processes using a variety of parameters.

CC Atorvastatin is used, in conjunction with dietary restriction, in the

CC management of hyperlipidaemia, including hypercholesterolaemia, mixed

CC dyslipidaemia and homozygous familial hypercholesterolaemia. The present

CC sequence encodes a nitrilase polypeptide obtained from an environmental

CC sample.

XX Sequence 1014 BP; 230 A; 304 C; 284 G; 196 T; 0 U; 0 Other;

Query Match 49.3%; Score 499.4; DB 12; Length 1014;  
 Best Local Similarity 68.3%; Pred. No. 1.1e-152;  
 Matches 692; Conservative 0; Mismatches 321; Indels 0; Gaps 0;

```

OY 2 TGAAGAAGCTATCAAGTGTGCGTGCAGACCCGCCGATCTACATGATTTGAGG 61
DB 2 TGAAGAAGCTATCAAGTGTGCGTGCAGACCTCCAGCTCTTCTCGACCTGAGG 61
OY 62 CCAAGGTGACAAACATTAAGTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 121
DB 62 CCAAGGTGACAAACATTAAGTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 121
OY 62 CCAAGGTGACAAACATTAAGTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 121
DB 62 CCAAGGTGACAAACATTAAGTATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 121
OY 122 TCGCTTTCCGGAAGCTTGTGATTCAGAGCTACCATAGTTTCTTGGCTTACTCAG 181
DB 122 TCGCTTTCCGGAAGCTTGTGATTCAGAGCTACCATAGTTTCTTGGCTTACTCAG 181
OY 122 TCGCTTTCCGGAAGCTTGTGATTCAGAGCTACCATAGTTTCTTGGCTTACTCAG 181
DB 122 TCGCTTTCCGGAAGCTTGTGATTCAGAGCTACCATAGTTTCTTGGCTTACTCAG 181
OY 182 CATGGCAATGCAATTTGTAGCGCAATACCATGAACTCATTTGGAGTTGGACCTTC 241
DB 182 CATGGCAATGCAATTTGTAGCGCAATACCATGAACTCATTTGGAGTTGGACCTTC 241
OY 242 AACCTAAGCCATTTTCAGATGACGCAAGGGGTTGGGAATCATGTCACCTGGGAGTGA 301
DB 242 AACCTAAGCCATTTTCAGATGACGCAAGGGGTTGGGAATCATGTCACCTGGGAGTGA 301
OY 302 GTGAACGGGTGCGGTGACACCTTTTACATGATGATTCATAGGCGATTAATGATGACA 361
DB 302 GTGAACGGGTGCGGTGACACCTTTTACATGATGATTCATAGGCGATTAATGATGACA 361
OY 362 CCAATGGGGGGGGGGGAGAGTTGAAACCTTCTTGTGAACTGATCTTGTGGCGGAG 421
DB 362 CCAATGGGGGGGGGGGAGAGTTGAAACCTTCTTGTGAACTGATCTTGTGGCGGAG 421
OY 422 GGGATGTTGATGCTGCTGAGAGCTGCTGTTGAGAGGCTGCTGCTTATGCT 481
DB 422 GGGATGTTGATGCTGCTGAGAGCTGCTGTTGAGAGGCTGCTGCTTATGCT 481
OY 482 GTTGGAGACACCTTCAACCGCTAACAAATACGCTTTGTATGACAAATGAGAGATTC 541
DB 482 GTTGGAGACACCTTCAACCGCTAACAAATACGCTTTGTATGACAAATGAGAGATTC 541
OY 542 ATTGTGGGCTTGGCGGAGCTTTAGCCTTTATCTTAATGCGGCGAAAGCCCTGGGGCTG 601
DB 542 ATTGTGGGCTTGGCGGAGCTTTAGCCTTTATCTTAATGCGGCGAAAGCCCTGGGGCTG 601
OY 542 ACTTGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 601
DB 542 ACTTGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 601
OY 602 ATGTCAATGTAGCGGCTCTGCAATCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 661
DB 602 ATGTCAATGTAGCGGCTCTGCAATCTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 661
OY 662 CGTGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 721
DB 662 CGTGTGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 721
OY 722 GCTTCTTCAAGCGCGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 781
DB 722 GCTTCTTCAAGCGCGCGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 781
OY 782 TCGGCGCTCTTGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 841
DB 782 TCGGCGCTCTTGGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 841
OY 842 GCATCTTGTCTAATAATGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 901
DB 842 GCATCTTGTCTAATAATGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 901
OY 902 GCTTCTAATAATGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 961
DB 902 GCTTCTAATAATGCGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 961
OY 962 CTTACGCTTTGGGTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1014
DB 962 CTTACGCTTTGGGTAAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 1014
  
```

RESULT 11  
 AD164300  
 ID AD164300 standard; DNA; 1014 BP.  
 XX  
 AC AD164300;  
 XX  
 XX 22-APR-2004 (first entry)  
 XX  
 DE DNA encoding nitrilase seq id 24.  
 XX  
 KM (R)-ethyl 4-cyano-3-hydroxybutyric acid; nitrile hydrolysis;  
 KM carboxylic acid; cyanohydrin moiety hydrolysis;  
 KM aminonitrile moiety hydrolysis; chiral alpha-hydroxy acid molecule;  
 KM chiral amino acid molecule; (S)-ethyl 4-cyano-3-hydroxybutyric acid;  
 KM (R)-mandelic acid; (S)-mandelic acid; (S)-phenyl lactic acid derivative;  
 KM (R)-phenyl lactic acid derivative; & enantiomeric excess;  
 KM & diastomeric excess; food additive; drug intermediate; ds; nitrilase;  
 KM gene.  
 XX  
 OS Unidentified.  
 XX  
 PN US2004014195-A1.  
 PD 22-JUN-2004.  
 XX  
 XX 15-MAY-2003; 2003US-00440523.  
 XX  
 XX 29-DEC-1999; 99US-0173609P.  
 XX 07-DEC-2000; 2000US-0254414P.  
 XX 28-DEC-2000; 2000US-00751299.  
 XX 21-JUN-2001; 2001US-0300189P.  
 XX 30-JUL-2001; 2001US-0309066P.  
 XX 22-JAN-2002; 2002US-0351336P.  
 XX 15-MAY-2002; 2002US-00146772.  
 XX 09-SEP-2002; 2002US-00241742.  
 XX  
 PA (DIVE-) DIVERSA CORP.  
 XX  
 XX Desantis G, Short JM, Burk MJ, Wong K, Farwell R, Chatman K;  
 P1 WPI; 2004-121569/12.  
 XX P-PSDB; AD164301.  
 DR  
 XX  
 XX Novel isolated or recombinant polypeptide having nitrilase activity,  
 PT useful in production of food additives.  
 PS  
 PS Claim 1; SEQ ID NO 47; 105pp; English.  
 XX  
 CC The invention describes an isolated or recombinant polypeptide (I)  
 CC comprising amino acids having a sequence at least 50 % identical to a  
 CC sequence (SI) available in electronic form (EC) from the following web  
 CC site ftp.segdata.uspto.gov/sequence.html?DocID=2004014195, or its  
 CC variants, having one or more mutations at residue 55 Lys, Gly or Glu, at  
 CC residue 60 glutamic acid, at residue 111 Ser, their combinations or  
 CC fragments. (I) is useful for: producing an (R)-ethyl 4-cyano-3-  
 CC hydroxybutyric acid; hydrolyzing a nitrile to a carboxylic acid;  
 CC chiral alpha-hydroxy acid molecule or an aminonitrile moiety; producing a  
 CC chiral alpha-hydroxy acid molecule or a chiral amino acid molecule;  
 CC producing an (S)-ethyl 4-cyano-3-hydroxybutyric acid; producing an (R)-  
 CC mandelic acid or (S)-mandelic acid; producing (S)-phenyl lactic acid  
 CC derivative or an (R)-phenyl lactic acid derivative; modifying a molecule;  
 CC and for identifying a modified compound. The inventive method is useful  
 CC for monitoring or determining & enantiomeric excess or & diastomeric  
 CC excess. (I) is useful in the production of food additives and drug  
 CC intermediates. This sequence encodes a nitrilase of the invention.  
 XX  
 SQ Sequence 1014 BP; 230 A; 304 C; 284 G; 196 T; 0 U; 0 Other;

Query Match 49.3%; Score 499.4; DB 12; Length 1014;  
 Best Local Similarity 68.3%; Pred. No. 1.1e-152;  
 Matches 692; Conservative 0; Mismatches 321; Indels 0; Gaps 0;

QY	2	TGAAGAAGCTTTCAAGAGTGCCTGCTGAGTGAACCGCCCGCATCTCAATGATTTGGAGG	61
Db	2	TGAAGAAGCAATCAAGAATGACTGTGTGCMAAGCTCCAGTCTTTCTGACCTGGACG	61
QY	62	CGACGGTGGACAAACCATTTGAGTTGATGGAAGACGACGCACTAATATGCTGCTGTA	121
Db	62	CCACAGTGGACAAAGACCGTGCCTGATTTGAGGAGGACGCCGTAAACGGGACAGCCTTA	121
QY	122	TGCGCTTTCCGGAAACTTGGATTTCCAGGCTACCCATGCTTTCTTTGGCTTGACTCACAG	181
Db	122	TGCGCTTTCCAAAGACTTGGATTTCCAGGCTACCCATGCTTTCTTTGGCTTGACTCACAG	181
QY	182	CATGGGCAATGCAATTTGTATGCGCCAAATACATAGAACTCATTTGGAGTTGGATGGCCCTC	241
Db	182	CCTGGGGGATGCAATTTGTTGCGCCGATACCAAGAAATCTACGTGTCTTGACAGCCCTC	241
QY	242	AAGCTTAAGCGCAATTCAGATGACGACCAAGCGGTGGGAATCATGTCCACCCTGGGAGTGA	301
Db	242	AGGCCAAAGGCATTCAGTGAAGCGCGCCAGGCGCCGGTATATACGTGCGCTTAGGGTACA	301
QY	302	GTTGAACGGGTCCGCTGCGACCCCTTTACATCAGTACAGTGTTCATATAGCGCATATATGTGACA	361
Db	302	GCGAAGCGGTGAGCGGAACCTCTCTAATAGGGGAGTGGCTCATTTGACGATTAAGGCGCAAA	361
QY	362	CCATTGGGGCCCGGCGAAAGTTGAACCTACTTTTGTGACGTAATTTTGTGGCGAAG	421
Db	362	CAGCTGGGGTGGCGCGAAGCTGAAACCAACCATATGAGCGAACCCCTTTGGTGAAG	421
QY	422	GGGATGTTCAATCGCTAGCGGTTTTGGAACGTCTGTGGAAAGCTGGGTGCTTATGCT	481
Db	422	GCGACGGATCATCCCTTTCCACTTTGACACACGTTGGGGGTGCTGGGCGACTGCT	481
QY	482	GTTGGGAGACCTTCAACCGCTTAACAATAATGCGCTTGTATGACAAATAAGAGATTC	541
Db	482	GTTGGGAACACTTACCAACTCTTTGAAATATGCGCTCTACGACAGAAAGAGAAATAC	541
QY	542	ATTGTCGAGCTTGGCGCGAGCTTTAGCGCTTATCTATGCGGCGAAGGCCCTGGGCGCTG	601
Db	542	ACTTGCGCGCTGTGGCTGACTTCAAGATCTACCGTGAAGGACAGAAAGTCTTTGACACAG	601
QY	602	ATGTCAATGTAGCGGCTCTCGAATTTATGCGGTTGAAGGGCAATGCTTGTGTAAGCT	661
Db	602	AAGTAAATGTCCAGACTTCTCGAATCTACGCGGTGAAGGGCAGTGTTTGTCTGCTT	661
QY	662	CGTGGCGGTGTTTCACAATCCATGATGATATGCTTTGTATCAGATGACGAAGGATG	721
Db	662	CTTGCGCGTCTGCTTGCGCAGAGATGATGAATATGCTTGCATGACGAAGGCAAGACA	721
QY	722	CGTTGCTTCTGCGCTGGTGTGACACTCAAGTATCAATAGGGGCTGATGTGTGACTTGG	781
Db	722	GCTTCTTCAAGCGCGGCGGCGGTACTCCGCAATTAATGCGTCCGATGGAAGGACGACTAG	781
QY	782	TGCGGCTCTTGGCCGAAAAAGAAAGGGATTTCTTACGCAAACTTTGATCTTGAAGTAC	841
Db	782	CGCGGCCCTTGGCGCAAAAAGAGAAAGGATTTCTTATGCACTCTGAGACCCCTGCGCTC	841
QY	842	GCATCCTTGTCTAAATATGCGGACAGACCTTGCTGCTATTTATCCGTCGCCGACATTAATC	901
Db	842	GAATCTATGCAAAAGACCGGACGTGATCCAGCCGGGCACTAATCTCAAGACGAGTCACTC	901
QY	902	GCTTGCTAATGATGCGACGCCCTTAATTAACGGTATGMAATGGAAGTGATCTTGTGTC	961
Db	902	GAGTCTGATCAATGCAAGTGCACATCAACGAGTGTGAGAGGTTGGAAGGAAATACGTG	961
QY	962	CTTACGCTTTGGGTAAAGCGTCTGAGACGGGTGCGCACTCCAAAGAAATTTGA	1014
Db	962	CATGCGCCCAAGGCTTTGAAGTTGAGGCGGCGCCCGGGTATCGAAAGCGATTTGA	1014

```

RESULT 12
ADC23790
ID      ADC23790 standard; DNA; 1014 BP
XX

```

AC		ADCC3790;
XX		
DT		18-DEC-2003 (first entry)
DE		DNA sequence (SeqID 57) encoding a nitrilase enzyme.
XX		
KW		gene; ds; nitrilase; nitrile; cyanohydrin; ammonia; biocatalyst; enantiomer; chiral medicine.
OS		Unidentified.
XX		
PN		WO2003000840-A2.
PD		
XX		03-JAN-2003.
PF		15-MAY-2002; 2002MO-USO15963.
XX		
PR		21-JUN-2001; 2001US-0300189P. 30-JUL-2001; 2001US-030906P. PR 22-JAN-2002; 2002US-035133CP.
PA		(DIVE-) DIVERSA CORP. (MADD/) MADDEN D.
PI	Madden M,	Desantis G, Chaplin JA, Weiner DP, Milan A, Chi E, Short JM, Burk M;
DR		WP1: 2003-201417/19. P-PsDB; ADCC23791.
PT		Noval nitrilase polypeptide, useful for making (R)- or (S)-ethyl-4-cyano- 3-hydroxybutyric acid or (R)- or (S)-mandelic acid or (S)- or (R)-phenyl lactic acid derivative and for producing pharmaceutical composition, and food additive.
PS	Claim 1; SEQ ID NO 57; 560bp; English.	
CC	This invention relates to nitrilases and the nucleic acids that encode	
CC	these enzymes thereof. Specifically, it refers to polypeptides that	
CC	exhibit nitrilase activity, i.e. the ability to directly hydrolyse	
CC	nitriles or cyanohydrins into their corresponding carboxylic acids and	
CC	ammonia. Nitrilases have commercial utility as biocatalysts for use in	
CC	the synthesis of enantioselectively pure aromatic and aliphatic amino	
CC	acids' as well as hydroxy acids, which are important for the development	
CC	of chiral medicines. Furthermore, the present invention describes	
CC	nitrilases, isolated from mesophilic microorganisms, that have improved	
CC	activity and stability at increased pH and temperature. They are also	
CC	inactive, efficient catalysts, have broad substrate specificity and	
CC	are capable of chiral differentiation. This polynucleotide is a DNA	
CC	sequence that encodes a nitrilase enzyme of the invention.	
SQ	Sequence 1014 BP; 228 A; 305 C; 285 G; 196 T; 0 U; 0 Other;	
Query Match	49.1%; Score 497.8; DB 10; Length 1014;	
Best Local Similarity	68.2%; Pred. No. 3.6e-152;	
Matches 691; Conservative 0; Mismatches 322; Indels 0; Gaps 0		
OY	2 TGAAAGAACCTATCAAGGTCGGCTCGTGCAGGCCGCCCATGTACTACATGATTGGAGC	61
Db	2 TGAAGAAGCAATCAAAGTAGCCCTGTGTGCAGAGCACGCTCCAGTCTTTCTTGACCCTGGACG	61
OY	CGACGGGGAACAACAATTGAGTTGATGGAAAGCAGCAGCTATAATGCTGTCGA	121
Db	62 CCACAGTGCACAAACCGTCGCCCCCTGATTAGAGAGGAGGCCCGTAACGGCGACGCTTAA	121
OY	122 TC GCCCTTCCGAAAACCTTGATTCCAGGCTACCCATGGTTTTCTTTGGCTTACTACACG	181
Db	122 TCGCCTTTCCAGAGACCTCGGATTCGAGGCTACCCATGGTTCTTTTGGGTGATCACCAG	181
OY	182 CATGGGCAATGCAATTTGTATGCGCAATTAACATGAGAACTATTGGAAGTGGATGGCCCTC	241
Db	182 CTGGGGGAGATGCAATTTGTGCGCGGAATACCAAGGAATCACTGATCTTGACAGCCCTC	241

Seq	Sequence	1014 BP; 228 A; 305 C; 285 G; 196 T; 0 U; 0 Other;
Query	Match	49.1%; Score 497.8; DB 10; Length 1014;
Best Local	Similarity	68.2%; Pred. No. 3.6e-152;
Matches	691; Conservative	0; Mismatches 322; Indels 0; Gaps 0
QY	2 TGAAGAAGCTATCAAGTCCCTGCGTGCAGCGCCCGCATCTACATGATTTGGAG	61
DB	2 TGAAGAAGCAATCAAGTAGCTGTGTGCAAGAGCTCCAGTCTTCTCGACCTGACG	61
QY	62 CGACGGTGCACAAACCATTTGATTTGATGGAAGACGACGACGTATATATGCTGTCTGA	121
DB	62 CCACAGTGCACAAACCGTGCCTCGATGTAGGAGGACGCCGTGATACGGCCACGCTTAA	121
QY	122 TCGCCTTTCCGGAACCTTGATTTCCAGGCTTACCATGGTTTCTTTGGCTTACTACACAG	181
DB	122 TCGCCTTTCCAGAACCTCGGATTTCCAGGCTTACCATGGTTTCTTTGGCTTACTACACAG	181
QY	182 CATGGCAATTCGATTTGTATGCGCAATTACATGGAATCACTATTGAGATTGGATGCCCTC	241
DB	182 CTTGGGGGATTCGATTTGTGTGCGGATTAACACGAGAACTCACTGGTCTCTGACAGCCCTC	241



Db 542 ACTTGCCGCGCTGAGCTTACAGCATCTACCGTCAAGCAGACAGAAAGTCTTGGACGAG 601  
Qy 602 ATGTCAATGTAAGCGGCTCTGCAATCTAATGCGTTGAAGGGAAGTCTTCCGTAAGCGT 661  
Db 602 AAGTAAATGTGGCAGCTTCTCGAATCTAGCGCGTGAAGGGCAGTGTGTTCTCGCTT 661  
Qy 662 CGTGGCGCTGCTTCAACAATCATGATGATGATCTTGTACATGATGACGAAGAAGCATG 721  
Db 662 CTGTGGCGCTGCTCTCTCGCAGATGATGAAATGCTCTGCACTGACGAAGCAAGCACA 721  
Qy 722 CGTGTCTTCTGCTGTGTGTGACATCACTGATCATAGAGGCTGTAGTGTGTGACTGG 781  
Db 722 GCCTTCTTCAAGCGCGCGCGGCTACTCCCGCATTCGCTCCGATGGCAGACGACTAG 781  
Qy 782 TCGCGCTCTTGGCCCAAAATGAAGGGTATCTCTAGGCAAACTTGTATCTTGAAGTAC 841  
Db 782 CGCGCCCTTGGCGCAAAACGAGGAAGTATCTTATAGCCACTGTGACCTTGGCGCTC 841  
Qy 842 GCATCTTGTCTTAAATGAGCGGCGGACGCTGTGTGATTTATCCGCTCCGACATTAATC 901  
Db 842 GAATCTATGCAAAAGCCGACGCTGATCCAGCCGGGCACTACTCCAGACGACGTCATCTC 901  
Qy 902 GCTTGTCTTATGATGACGAGCCCTTAATTAACCGGTAGTGAATGGAAGTATCTTGTGTC 961  
Db 902 GAGTGTCTGATCAATGCGAGTGTCCATTCAGCCAGTGTGAGAGTTGAGCGGAAATATCCTG 961  
Qy 962 CTTACGCTTTGGGTAAAGCGTGTGAGCGGGTGGCCAACTCGAAGAAATTGCA 1014  
Db 962 CATGCGCCAAAGGCTTTGAAGTTGAGCGGCGCCCGGGGTACGAGGCGCATTTGA 1014

RESULT 14  
ADG93592  
ID ADG93592 standard; DNA; 1014 BP.  
AC ADG93592;  
DT 11-MAR-2004 (first entry)  
DE Nitriase enzyme gene sequence Segid57.  
KW nitriase; nitrite; carboxylic acid; chemical process; pH; temperature;  
KW enantioselective transformation; gene; ds.  
OS Unidentified.  
PN MO2003097810-A2.  
PD 27-NOV-2003.  
PF 15-MAY-2003; 2003MO-US015712.  
PR 15-MAY-2002; 2002US-00146772.  
PR 09-SEP-2002; 2002US-00241742.  
XX (DIVE-) DIVERSA CORP.  
PI Desantis G, Short JM, Burk M, Wong K, Farwell R, Chatman K;  
DR MPI; 2004-090637/09.  
DR P-PSDB; ADG93593.  
XX  
XX  
PS Claim 44; SEQ ID NO 57; 295bp; English.  
CC This invention is related to a novel isolated or recombinant nucleic acid encoding a protein having nitriase activity. Nitriase's are capable of converting nitrite's directly to carboxylic acids and have great potential for use in industrial chemical processes. The isolated CC nitriase proteins of the invention have increased activity and stability at increased pH and temperature when compared to those conventionally

CC used. In addition, the nucleic acid of the invention is useful for CC screening enantioselective transformation. The present sequence is that CC of a DNA sequence which encodes a nitriase enzyme of the invention.  
SQ Sequence 1014 BP; 228 A; 305 C; 285 G; 196 T; 0 U; 0 Other;  
Query Match 49.1%; Score 497.8; DB 12; Length 1014;  
Best Local Similarity 68.2%; Pred. No. 3.6e-152;  
Matches 691; Conservative 0; Mismatches 322; Indels 0; Gaps 0;  
Qy 2 TGAAGAAAGTATCAAGTTCGCTGCTGCAAGCCGCCGATCTTACATGATTTGAGG 61  
Db 2 TGAAGAAAGTATCAAGTTCGCTGCTGCAAGCCGCCGATCTTACATGATTTGAGG 61  
Qy 62 CGACGTGTGAAGAAACCTTGAAGTTGAGAAAGACGACGTAATATGTCGTGTA 121  
Db 62 CGACGTGTGAAGAAACCTTGAAGTTGAGAAAGACGACGTAATATGTCGTGTA 121  
Qy 122 TCGCCTTCCGGAACCTTGAATCCAGGCTACCCATGCTTCTTGGCTTGACGACG 181  
Db 122 TCGCCTTCCGGAACCTTGAATCCAGGCTACCCATGCTTCTTGGCTTGACGACG 181  
Qy 182 CATGGCAATGCAATTTGTACGCAATACATGAGAACTATGAGTTGATGAGCCCTC 241  
Db 182 CTGGGGGATGCAATTTGTGCGCGATACACAGAACTCACTGTCTTCCGACGACCTTC 241  
Qy 242 AAGCTAAGCGCATTTCAATGATGACGACGAGCGGTTGGGAATCATGTCAACCTTGGGAGTA 301  
Db 242 AAGCGCAACGGATCATGATGAGCGCGCGCGCGGTATATACCTTCGCGTGGGTACA 301  
Qy 302 GTGAACGGGTGTGGTGGACCCCTTTATCATGATGATGATGATGATGATGATGATG 361  
Db 302 GTGAACGGGTGTGGTGGACCCCTTTATCATGATGATGATGATGATGATGATGATG 361  
Qy 362 CCATTGGGGCCCGGGAAGTTGAAACCTTATTTGTGAAAGTACTTTTGTTCGCGAAG 421  
Db 362 CAGCTGGGTGCGCGGGAAGTGTGAAACCAATGATGAGGGAACCTCTTGGTGAAG 421  
Qy 422 GGGATGTTTATCGGTACCGGTTTTCAGACGTCGTGTGGAAGGCTGGGTGCTTATGCT 481  
Db 422 GCGACGATCATCTCTTTCACCTTTCAGACGTCGTGTGGAAGGCTGGGTGCTTATGCT 481  
Qy 482 GTTGGAGACCTTCAACCGCTTAAACAAATAGCGTTGTATGACCAAAATGAGAGATTG 541  
Db 482 GTTGGAGACCTTCAACCGCTTAAACAAATAGCGTTGTATGACCAAAATGAGAGATTG 541  
Qy 542 ATTGTGCGGCTTGGCCGAGCTTTAGCCTTTATTCCTAATGCGCGGAACCCCTGCGGCTG 601  
Db 542 ACTTGCCGCGCTTGGCCGAGCTTTAGCCTTTATTCCTAATGCGCGGAACCCCTGCGGCTG 601  
Qy 602 ATGTCAATGTAAGCGGCTCTGCAATCTAATGCGGTTGAAGGCAATGCTTCTGTAAGCGT 661  
Db 602 AAGTAAATGTGGCAGCTTCTCGAATCTAGCGCGGTGAAGGCGCAGTGTGTTCTCGCTT 661  
Qy 662 CGTGGCGCTGCTTCAACAATCATGATGATGATCTTGTACATGATGACGAAGAAGCATG 721  
Db 662 CTGTGGCGCTGCTCTCTCGCAGATGATGAAATGCTCTGACATGACGAAGAAGCACA 721  
Qy 722 CGTGTCTTCTGCTGTGTGTGACATCACTGATCATAGAGGCTGTAGTGTGTGACTGG 781  
Db 722 GCCTTCTTCAAGCGCGCGGCTACTCCCGATTAATGCTTCCGATGGCAGCAGCTAG 781  
Qy 782 TCGCGCTCTTGGCCCAAAATGAAGGGTATCTCTAGGCAAACTTGTATCTTGAAGTAC 841  
Db 782 CGCGCCCTTGGCGCAAAACGAGGAAGTATCTTATAGCCACTGTGACCTTGGCGCTC 841  
Qy 842 GCATCTTGTCTTAAATGAGCGGCGGACGCTGTGTGATTTATCCGCTCCGACATTAATC 901  
Db 842 GAATCTATGCAAAAGCCGACGCTGATCCAGCCGGGCACTACTCCAGACGACGTCATCTC 901  
Qy 902 GCTTGTCTTATGATGACGAGCCCTTAATTAACCGGTAGTGAATGGAAGTATCTTGTGTC 961  
Db 902 GAGTGTCTGATCAATGCGAGTGTCCAAATCAGCAGTGTGAGAGTTGAGCGGAAATATCCTG 961







GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 26, 2006, 22:54:39 ; Search time 209.215 Seconds  
(without alignments)  
8615.305 Million cell updates/sec

Title: US-09-751-299-3

Perfect score: 1014

Sequence: 1 atgaagaagcatcaaggt.....cgcaactcgagaatttga 1014

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 1303057 seqs, 888780828 residues

Total number of hits satisfying chosen parameters: 2606114

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-Processing: Minimum Match 0%

Maximum Match 100%

Database : Issued Patents NA:\*

1: /cgn2\_6/prodata/1/ina/1/COMB.seq:\*  
2: /cgn2\_6/prodata/1/ina/5/COMB.seq:\*  
3: /cgn2\_6/prodata/1/ina/6A/COMB.seq:\*  
4: /cgn2\_6/prodata/1/ina/6B/COMB.seq:\*  
5: /cgn2\_6/prodata/1/ina/H/COMB.seq:\*  
6: /cgn2\_6/prodata/1/ina/ECTUS/COMB.seq:\*  
7: /cgn2\_6/prodata/1/ina/PP/COMB.seq:\*  
8: /cgn2\_6/prodata/1/ina/RE/COMB.seq:\*  
9: /cgn2\_6/prodata/1/ina/backfilest1.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	283	27.9	1035	2	US-08-690-493-2
2	283	27.9	1200	2	US-08-690-493-3
3	248.6	24.5	1071	3	US-09-806-876A-1
4	218	21.5	1110	3	US-09-823-373-4
5	218	21.5	1110	3	US-09-823-373-13
6	218	21.5	1776	3	US-09-823-373-15
7	216.8	21.4	1110	3	US-09-823-373-16
8	132.4	13.1	1194	2	US-08-447-702-4
9	132.4	13.1	1194	2	US-08-465-615-4
10	84	8.3	385	3	US-09-823-373-3
11	36.6	3.6	601	3	US-09-949-016-66400
12	36.6	3.6	110243	3	US-09-949-016-13698
13	35.4	3.5	90	3	US-09-823-373-30
14	35	3.5	90	3	US-09-823-373-27
15	34.8	3.4	90	3	US-09-823-373-32
16	34.2	3.4	601	3	US-09-949-016-60707
17	34.2	3.4	605	3	US-09-270-767-546
18	34.2	3.4	605	3	US-09-270-767-15828
19	34.2	3.4	1434	3	US-09-491-772-2
20	34.2	3.4	8580	3	US-09-491-772-1
21	34.2	3.4	283538	3	US-09-949-016-13506
22	33.8	3.3	601	3	US-09-949-016-60708
23	33	3.3	1401	4	US-09-605-703B-543
24	32.8	3.2	1068	3	US-09-949-016-3762

25	32.8	3.2	1115	3	US-09-949-016-539	Sequence 539, App
26	32.8	3.2	2088	3	US-09-023-655-340	Sequence 340, App
27	32.6	3.2	336	3	US-09-643-597-286	Sequence 286, App
28	32.6	3.2	336	3	US-09-480-884A-286	Sequence 286, App
29	32.6	3.2	336	3	US-09-542-615A-286	Sequence 286, App
30	32.6	3.2	336	3	US-09-606-421B-286	Sequence 286, App
31	32.6	3.2	336	3	US-09-630-940B-286	Sequence 286, App
32	32.6	3.2	336	3	US-10-007-700-286	Sequence 286, App
33	32.6	3.2	2835	3	US-09-134-001C-1515	Sequence 1515, App
34	32.6	3.2	3561	3	US-09-134-001C-1685	Sequence 1685, App
35	32.6	3.2	6414	3	US-09-134-001C-1626	Sequence 1626, App
36	32.4	3.2	561	3	US-09-902-540-6607	Sequence 6607, App
37	32.4	3.2	2260	3	US-09-902-540-502	Sequence 502, App
38	32.2	3.2	505	3	US-09-621-976-15639	Sequence 15639, App
39	32	3.2	2652	2	US-08-318-831-1	Sequence 1, App
40	31.8	3.1	894	3	US-09-489-039A-5631	Sequence 5631, App
41	31.8	3.1	1930	3	US-09-270-767-9205	Sequence 9205, App
42	31.6	3.1	1930	3	US-09-270-767-24487	Sequence 24487, App
43	31.6	3.1	29653	3	US-09-949-016-13442	Sequence 13442, App
44	31.6	3.1	37304	3	US-09-949-016-11862	Sequence 11862, App
45	31.6	3.1	481115	3	US-09-949-016-11940	Sequence 11940, App

## ALIGNMENTS

RESULT 1  
US-08-690-493-2  
; Sequence 2, Application US/08690493  
; Patent No. 5872000  
; GENERAL INFORMATION:  
; APPLICANT: Yu, Fujio  
; TITLE OF INVENTION: No. 5872000e1 Nitrilase Gene  
; NUMBER OF SEQUENCES: 5  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Steinberg, Raskin & Davidson, P.C.  
; STREET: 1140 Avenue of the Americas  
; CITY: New York  
; STATE: New York  
; COUNTRY: U.S.A.  
; ZIP: 10036  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: MS-DOS Editor  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/690,493  
; FILING DATE: 31 JUL-1996  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 213061/1995  
; FILING DATE: 31-JUL-1995  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Davidson, Clifford M  
; REGISTRATION NUMBER: 32,728  
; REFERENCE/DOCKET NUMBER: 3821005  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (212) 768-3800  
; TELEFAX: (212) 382-2124  
; INFORMATION FOR SEQ ID NO: 2:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 1035 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: double  
; TOPOLOGY: linear  
; MOLECULE TYPE: genomic DNA  
; ORIGINAL SOURCE:  
; ORGANISM: Gordona terrae  
; STRAIN: MA-1  
; CELL TYPE: unicellular organism  
; US-08-690-493-2  
Query Match 27.9%; Score 283; DB 2; Length 1035;

Best Local Similarity 57.5%; Pred. No. 1.7e-82;  
Matches 508; Conservative 0; Mismatches 375; Indels 0; Gaps 0;

```
QY 14 TCAAGTGGCTGCTGGAAGCCCGCATCTACATGATTTGAGGCGACGGTGACA 73
Db 26 TCAAGGACGCGGTGACCCAGGCGGAGCGGTGTGATTCGCGCCACCGTGACA 85
QY 74 AAACCATTTAGTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 133
Db 86 AGACCATTTGGCGGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 145
QY 134 AAACCTTGATTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 193
Db 146 AGACCTTGATTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 205
QY 194 AATTGTGACGCAATATACATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 253
Db 206 AGTACTTATCGGTGATTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 265
QY 254 TTTGAGATGACAGCCAGAGCGGTGGGAAATCATGTGACCCCTGGGAGATGAGTA 313
Db 266 TCAGGAGAGCGGACGAGAAAGACATCGCATCACCATGGGATTTAGTAGCGCGGTC 325
QY 314 GTGGACCCCTTTACATCAGTCAAGTGTTCATAGGAGATTAATGTGACACATGG 373
Db 326 ATGTTGCTGTGACATGGGCGGAGGCGGTATCGAGCGGAGCGGAGTGTGACGAC 385
QY 374 GCGGAAAGTTGAAACCTATTGTTGTAACGTAATTTGTTGCGGAGAGGAGAGTTCAT 433
Db 386 GCGGAAAGTTGAAACCTATTGTTGTAACGTAATTTGTTGCGGAGAGGAGTTCAT 445
QY 434 CGTACGCGTTTTCGAGAGCTGTGTTGGAAGCGTGGGCTTATGCTGTTGGAGACC 493
Db 446 ATCTGTGTGTGAGACAGACCAAGTCTCGGCGGAGTGGGTGCTGTGTTGGAGACC 505
QY 494 TTTCAACCCCTTAAACAAATATGCTTTGTATGCAACAATGAGAGATTCATTTGG 553
Db 506 TCGAGCCGTGACCAAGTACGCGCATGTACTGCAACAGAGAGATTCATTCGCGCAT 565
QY 554 GCGCGAGCTTTAGCTTTATCTTAATGCGGCGAGAAAGCCCTGGGCGCTGATGTCAT 613
Db 566 GCGCGAGCTTTAGCTTTATCTTAATGCGGCGAGAAAGCCCTGGGCGCTGATGTCAT 625
QY 614 GCGCGCTCTGCAATCTATGCGGCTTGAAGGCAATGCTTGATGACGCTGTGCTG 673
Db 626 GCGCGCTCTGCAATCTATGCGGCTTGAAGGCAATGCTTGATGACGCTGTGCTG 685
QY 674 TTTGCAATTCATGATGATGATGATGATGATGATGATGATGATGATGATGATG 733
Db 686 TCGGAGATGCAAGTGTGGAGGCGTTTGCGGATACCGAGAGAGAGAGAGAGAGAG 745
QY 734 CTGTGTGAGACACTCACGTATCATAGGAGCTGATGTGTGATCTTGTGCGGCTT 793
Db 746 AAGGAGGCGGATATCCCGTATCTACGGATCCGAGCGGTGTTCACTCGGAGAAC 805
QY 794 CCGAAATGAGAGAGGATTTCTTACGCAAACTTGATCTGTGAGATGACATCTTGT 853
Db 806 CGCCCAATGAGAGAGGATTTCTTACGCAAACTTGATCTGTGAGATGACATCTTGT 865
QY 854 AAATGGCGGAGAGAGCTGTGTGATTAATCCGTCGCGAGCAT 896
Db 866 AGAACCCGCGGAGAGAGCTGTGTGATTAATCCGTCGCGAGCAT 908
```

## RESULT 2

US-08-690-493-3

; Sequence 3, Application US/08690493

; Patent No. 5872000

; GENERAL INFORMATION:

; APPLICANT: Yu, Fujio

; TITLE OF INVENTION: No. 5872000e1 Nitrlilase Gene

; NUMBER OF SEQUENCES: 5

; CORRESPONDENCE ADDRESS:

ADDRESSEE: Steinberg, Raskin & Davidson, P.C.  
STREET: 1140 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: MS-DOS Editor

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/690,493

FILING DATE: 31 JUL-1996

PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP 213061/1995

FILING DATE: 31-JUL-1995

ATTORNEY/AGENT INFORMATION:

NAME: Davidson, Clifford M

REGISTRATION NUMBER: 32,728

REFERENCE/DOCKET NUMBER: 3821005

TELECOMMUNICATION INFORMATION:

TELEPHONE: (212) 768-3800

TELEFAX: (212) 382-2124

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 1200 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: genomic DNA

ORIGINAL SOURCE:

ORGANISM: Gordona terrae

STRAIN: MA-1

CELL TYPE: unicellular organism

US-08-690-493-3

Query Match 27.9%; Score 283; DB 2; Length 1200;

Best Local Similarity 57.5%; Pred. No. 1.9e-82;

Matches 508; Conservative 0; Mismatches 375; Indels 0; Gaps 0;

```
QY 14 TCAAGTGGCTGCTGGAAGCCCGCATCTACATGATTTGAGGCGACGGTGACA 73
Db 120 TCAAGGACGCGGTGACCCAGGCGGAGCGGTGTGATTCGCGCCACCGTGACA 179
QY 74 AAACCATTTAGTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 133
Db 180 AGACCATTTGGCGGTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 239
QY 134 AAACCTTGATTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 193
Db 240 AGACCTTGATTTGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 299
QY 194 AATTGTGACGCAATATACATGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 253
Db 300 AGTACTTATCGGTGATTCGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 359
QY 254 TTTGAGATGACAGCCAGAGCGGTGGGAAATCATGTGACCCCTGGGAGATGAGTA 313
Db 360 TCAGGAGAGCGGACGAGAAAGACATCGCATCACCATGGGATTTAGTAGCGCGGTC 419
QY 314 GTGGACCCCTTTACATCAGTCAAGTGTTCATAGGCGATTAATGTGACACATTTGG 373
Db 420 ATGTTGCTGTGATGAGGCGGAGGCGGTGATGAGGCGGTGATGAGGCGGTGAT 479
QY 374 GCGGAAAGTTGAAACCTATTGTTGTAACGTAATTTGTTGCGGAGAGGAGAGTTCAT 433
Db 480 GCGGAAAGTTGAAACCTATTGTTGTAACGTAATTTGTTGCGGAGAGGAGAGTTCAT 539
QY 434 CGTACGCGTTTTCGAGAGCTGTGTTGGAAGCGTGGGTGCTTATGCTGTTGGAGACC 493
Db 540 ATCTGTGTGAGACAGACCAAGTCTCGGCGAGTGGGTGCTGCTGCTGTTGGAGACC 599
```



TYPE: DNA  
ORGANISM: Acidovorax facilis  
US-09-823-373-4

Query Match 21.5%; Score 218; DB 3; Length 1110;

Best Local Similarity 52.0%; Pred. No. 4,66-61; Mismatches 450; Indels 0; Gaps 0;

Matches 488; Conservative 0; Mismatches 450; Indels 0; Gaps 0;

```
22 GCTGCGTGAAGAGCCGCCGATCTACATGATTTGAGGCGACGGTGACAAAACCAT 81
31 GCAACCGTTGAGGAGAGCCGGTATGCTCGACGACGCAACGATGACAAAGTCATC 90
82 GAGTTGATGAAAGAGCAGCATATATGCTGCTGATGCTCTTCCGGAACCTGG 141
91 GGCATCATGAAAGAGCTGCCAAAAGGCGGAGATGATGCTTCTCCGAAATTC 150
142 ATTCAGGCTACCCATGCTTTCTTTGGCTTGACCTACACAGATGGGCAATCTTGT 201
151 ATTCGGGCTACCCCTATGCGGCGTGGCTCGGCGACGTAAGTACAGCTTAAGT 210
202 CGCCAAATACATGAGACTCATTTGAGTTGATGGCCCTCAAGCTAAGCCATTGAT 261
211 TCACGCTATCAGAAATTCGTTGAGCTAGGTGACGACCTGATCGCTCCAGCTG 270
262 GCAAGCAAGCGGTTGGGAATCATGCTCAACCTGGGAGTGAAGAGGAGTGGGAC 321
271 GCCGCGCGCGCAACAAATGCACTGCTCATGAGGCTATTGAGAGCGGAAAGCG 330
322 CTTTACATGATGATGCTGTTGATGAGCGATATGCTGACACCATTTGGGCGCG 381
331 CGCTATCGAGCGAGGTGTTGATCGACGAGCGTGGCGAGATGTTGCCAATGCG 390
392 TTGAAACCTACTTTTGTGAAAGTACTTTGTTGGCGGAAAGGAGTGTTCATCG 441
391 CTGAAGCCCAACAAGTTGAGGTGACATCTACGCGAAGGCAAGCCGATTTCTC 450
442 GTTTTCGAGAGCTGTTGAGAGGCTGGGCTGATGCTGTTGGAGACACTTCAAC 501
451 ACGCAGACTTGGCGTTGAGAGCGCTGGTGGATTTGAATGCTGGGAATTTCCAA 510
502 CTAAACAAATACGCTTTGTATGCAACAAATGAAGATTCATTTGCGGCTTTGCG 561
511 CTCACAGATTCATGATGTAAGCTCGGTGAGAGCTCAAGTTGCAATCGGCGCG 570
562 TTTAGCCTTTATCTTAATGCGCGGAAAGCCCTGGGCGCTGATGTAATGTAAG 621
571 ATGTCCTCTTTCAGCCGGATGTTTCCAACTGAGATCGAAGCAACGCGATCAC 630
622 CGAATCTATGCGGTTGAGAGGCAATGCTGTAAGCTGATGCGCTGTTTCAACA 681
631 CGCTGTAAGCAATGAAAGCCAACTTTGCTGCTTGTGCTGACGAGGTATCGAAC 690
682 TCCATGATGATGATGCTTTGTACAGATGCAAGAAAGCATGCTGCTTGTGCTGT 741
691 ACGCGGATGAAAGCTTCTGCTCAACGACGACGCGCACTGTGCGGCAAGATGT 750
742 GGAACATCAAGTATCATAGGCGCTGATGAGCTTGTGCGCGCTTTGCCGAAAT 801
751 GCGTGGGCGCGCATTTACGCGCGGATGGAAGGAGCTTGCAGGCTTGGCGAAG 810
802 GAAGAGGATATCTTACGCAAACTTGATCTGAGTACGCAATCTTGAATGGGCG 861
811 GCTGAGGGGATCTTGAAGCAGAGATGATCTGAGACAGATTTCTGTGCGAAGGT 870
862 GCAAGACCTGCTGCTGATTTATTCCTGCTCCGACATTAATGCTGTAATGCGAC 921
871 GCGGATCCGCTGCGGAGCATTTGCGCGCTGACGCTGCTGCTGCAATTCGAC 930
922 CTTAAATACCGGTAAGTGAATTTGAAGGATCTTCCG 959
931 AATCATAGCCAGTTATCGCATCGGATTTGACGGTGC 968
```

RESULT 5  
US-09-823-373-13

Sequence 13, Application US/09823373

Patent No. 6870038

GENERAL INFORMATION:

APPLICANT: Chauban, Sarita

APPLICANT: DiCosimo, Robert

APPLICANT: Payne, Mark

APPLICANT: Gavan, John

APPLICANT: Fallon, Robert

TITLE OF INVENTION: Isolation and Expression of a Gene for Nitrilase from

TITLE OF INVENTION: Acidovorax Facilis 72W

FILE REFERENCE: BC-1032 US NA

CURRENT APPLICATION NUMBER: US/09/823,373

PRIOR FILING DATE: 2001-03-30

PRIOR APPLICATION NUMBER: 60/193,707

PRIOR FILING DATE: 2000-03-31

NUMBER OF SEQ ID NOS: 32

SOFTWARE: Microsoft Office 97

SEQ ID NO 13

LENGTH: 1110

TYPE: DNA

ORGANISM: Acidovorax facilis

US-09-823-373-13

Query Match 21.5%; Score 218; DB 3; Length 1110;

Best Local Similarity 52.0%; Pred. No. 4,66-61; Mismatches 450; Indels 0; Gaps 0;

Matches 488; Conservative 0; Mismatches 450; Indels 0; Gaps 0;

```
22 GCTGCGTGAAGAGCCGCCGATCTACATGATTTGAGGCGACGGTGACAAAACCAT 81
31 GCAACCGTTGAGGAGAGCCGGTATGCTCGACGACGCAACGATGACAAAGTCATC 90
82 GAGTTGATGAAAGAGCAGCATATATGCTGCTGATGCTCTTCCGGAACCTGG 141
91 GGCATCATGAAAGAGCTGCCAAAAGGCGGAGATGATGCTTCTCCGAAATTC 150
142 ATTCAGGCTACCCATGCTTTCTTTGGCTTGACCTACACAGATGGGCAATGTA 201
151 ATTCGGGCTACCCCTATGCGGCGTGGCTCGGCGACGTAAGTACAGCTTAAGT 210
202 CGCCAAATACATGAGAACTCATTTGAGTTGATGGCCCTCAAGCTAAGCCATTG 261
211 TCACGCTATCAGAAATTCGTTGAGCTAGGTGACGACCTGATCGCTCCAGCTG 270
262 GCAAGCAAGCGGTTGGGAATCATGCTCAACCTGGGAGTGAAGAGGAGTGGGAC 321
271 GCCGCGCGCGCAACAAATGCACTGCTCATGAGGCTATTGAGAGCGGAAAGCG 330
322 CTTTACATGATGATGCTGTTGATGAGCGATATGCTGACACCATTTGGGCGCG 381
331 CGCTATCGAGCGAGGTGTTGATCGAGAGCGGTGGCGAGATGTTGCCAATGCG 390
392 TTGAAACCTACTTTTGTGAAAGTACTTTGTTGGCGGAAAGGAGTGTTCATCG 441
391 CTGAAGCCCAACAAGTTGAGGTGACATCTACGCGAAGGCAAGCCGATTTCTC 450
442 GTTTTCGAGAGCTGTTGAGAGGCTGGGCTGATGCTGTTGGAGACACTTCAAC 501
451 ACGCAGACTTGGCGTTGAGAGCGCTGGTGGATTTGAATGCTGGGAATTTCCAA 510
502 CTAAACAAATACGCTTTGTATGCAACAAATGAAGATTCATTTGCGGCTTTGCG 561
511 CTCACAGATTCATGATGTAAGCTCGGTGAGAGCTCAAGTTGCAATCGGCGCG 570
562 TTTAGCCTTTATCTTAATGCGCGGAAAGCCCTGGGCGCTGATGTAATGTAAG 621
571 ATGTCCTCTTTCAGCCGGATGTTTCCAACTGAGATCGAAGCCAGCGATCAC 630
622 CGAATCTATGCGGTTGAGAGGCAATGCTGTAAGCTGATGCGCTGTTTCAACA 681
631 CGCTGTAAGCAATGAAAGCCAACTTTGCTGCTTGTGCTGACGAGGTATCGAAC 690
```

QY 682 TCATGATGATATGCTTTGTACAGATGACGAAAAAGCATGCTGCTTCTGCGTGTGT 741  
DB 691 AGCGGATGGAACGTTCTGCTCAACGACGACGCGGCACTGTGTGCGGAAGATGT 750  
QY 742 GGAACATCACTGATCATAGAGGCTGTAGTGTGTGACTGTGTGCGCTCTTGTCCGAAAT 801  
DB 751 GCTGTGGGCGCGCATTTTACGCGCCCGATGAGACGAGCTTGCAGAACGCTCTGCGGAAAGAT 810  
QY 802 GAAAGAGGATTTCTCAACGAAACCTGTATCTCGAGTACGATCTTGTCTAAATAGCG 861  
DB 811 GCTGTAGGGGATCTTGACGACAGATGATCTGTGAGCAGATTTCTGTGCGAAGGCTGGA 870  
QY 882 GGAAGACCTGCTGTGATATTCCGCTCCGACATTAATCTGCTGTGTAATAGTGCAGC 921  
DB 871 GCGGATCCGGGTGGGCACTATTGCGGCGCTGACGCTGTCTGCTCAAGTTCAACCCGCGC 930  
QY 922 CCTAAATTAACCGGTAGTTGAAATTGAAAGGTGATCTTTCG 959  
DB 931 AATCATAGCGCAGTTTCATCGATCGGCAATGACGGTCTG 968

RESULT 6  
US-09-823-373-15  
; Sequence 15, Application US/09823373  
; Patent No. 6870038  
; GENERAL INFORMATION:  
; APPLICANT: Chaubhan, Sarita  
; APPLICANT: Diccolino, Robert  
; APPLICANT: Payne, Mark  
; APPLICANT: Gavagan, John  
; APPLICANT: Fallon, Robert  
; TITLE OF INVENTION: Isolation and Expression of a Gene for Nitrilase from  
; FILE REFERENCE: BC-1032 US NA  
; CURRENT APPLICATION NUMBER: US/09/823,373  
; CURRENT FILING DATE: 2001-03-30  
; PRIOR APPLICATION NUMBER: 60/193,707  
; NUMBER OF SEQ ID NOS: 32  
; SOFTWARE: Microsoft Office 97  
; SEQ ID NO 15  
; LENGTH: 1776  
; TYPE: DNA  
; ORGANISM: Acidovorax delafieldii  
US-09-823-373-15

Query Match 21.5%; Score 218; DB 3; Length 1776;  
Best Local Similarity 52.0%; Pred. No. 6e-61;  
Matches 488; Conservative 0; Mismatches 450; Indels 0; Gaps 0;  
QY 22 GCTTGGTTCGAAGCCGCCGATCTACATGATTTGGAGCGACGGTGGCAAAACATT 81  
DB 362 GCAACCGTTCAAGCAGAGCCGGTATGCTCGACGACGACGCAACATCGAACATCGATC 421  
QY 82 GAGTGTGAGAGAGCAGCAGTAATATGCTCGTGTGATGCTTCCGGAACCTGG 141  
DB 422 GGCATCATGAAAGAGCTCCCAAAAGGCGCGAGTGTGATGCTTCCGGAAGTATTC 481  
QY 142 ATTCAGGCTACCACTGTTCTTTGGCTTGAATCAACGACATGGGCAATGCAATTTGTA 201  
DB 482 ATTCGGGGTACCCATATTGGGCGGTGCTCGGCGACGTGAATACAGCTTAAGCTTACT 541  
QY 202 CGCCAAATACATGAAACTGATTTGAGTTGATGCGCTCAAGCTTAAGCGCATTTCAAT 261  
DB 542 TCACCTATCAAGAAATTCGTTGAGGTGAGTGAACCGTATGCTGCTCCAGCTG 601  
QY 262 GGAAGCAAGCGGTGGAGTCAATGTCACCCCTGGGATAGAGAACGGTCTGATGACCC 321  
DB 602 GCGGCGCGCGCAAAATTCGACTGCTCATGAGGTATTTGAGAGGAAACCGGATGCG 661  
QY 322 CTTCATCATGATGATGTTTCATAGCGATAATGATGACACCATTTGGGCGCGGAAG 381  
DB 662 CGCTATCTGAGCGAGGTTCATTCAGCAGCGGTGGGGAATGTTGCCAATGGGCGAAG 721

QY 382 TTGAACCTACTTTTGTGAACGATCTTGTTCGGGAAGGAGATGTTTCATCGCTACCG 441  
DB 722 CTGAAGCCCAACACATTTGAGCGTACATCTACCGCGAAGCAACGAAACGATTTCTTC 781  
QY 442 GTTTTCGAGACGCTGTGTGAAGGCTGGGTGCTTATGCTGTGTGAGACACCTTCAACCG 501  
DB 782 ACGACGACCTTCGCGTTCGACGCGTCCGTGGAATGAACCTGCTGGGAACATTTCAACCG 841  
QY 502 CTACAAATATACGCTTTGTATGACAAATGAAAGATTTCAATGTGCGGCTTGGCGGAGC 561  
DB 842 CTCAGCAAGTTCATGATATACAGCTCGGTGAGCAGGTCCACGTTGCAATCGTGGCGGCG 901  
QY 562 TTTAGCCTTTATCCATATGCGCGAAGCCGCGGGGCGCTGATGTCATGTAAGTGAAGCTCT 621  
DB 902 ATGTCCCTCTTCACCGCGAATGTTTCCAACTGACATGACAAACGACGAGGTACCC 961  
QY 622 CGAATCTATGCGGTGGAAGGCAATGCTTGTACTAGCGTGTGTGCGCTGTTTCAACA 681  
DB 962 CGCTGTGATGCAATGGAAGCGCAACCTTTGTGCTTGTCTGACGCAAGTATGCGGACCT 1021  
QY 682 TCATGATGATATGCTTTGTACAGATGACGAAAAAGCATGCGTGTCTTGTGCTGTGTGT 741  
DB 1022 AGCGGATGCAAACTTCTGCTCAACGACGACGCGCACTGTTCGCGAAGATGT 1081  
QY 742 GGAACCTGATCATATGAGGCGCTGATGATGTGTGCTTGTGTGCGGCTTGTGCGAAT 801  
DB 1082 GCTGTGGGCGCCATTTACGCGCCGATGAAAGCGAGCTTGTGAAAGCTTGTGCGGAAGT 1141  
QY 802 GAAAGGATTTCTTCAACGCAACCTTGTATCTGAGTACACATCTTGTCTTAAATGAGCG 861  
DB 1142 GCTGAGGGGATCTTGTACGACGAGATGATGATCTGACGAGATTTGCTGCGAAGCTGGA 1201  
QY 862 GCAAGCCTGCTGTGATTTATTTCCGTCGCAACATTAATCTGCTGTAATGATGACAGC 921  
DB 1202 GCGGATCCGGTGGGCACTATTCGCGGCTGACGCTGTGTGTGCTGCAATGACCGCGC 1261  
QY 922 CCTAAATTAACCGGTAGTTGAAATTGAAAGGTGATCTTTCG 959  
DB 1262 AATCATAGCGCAGTTTCATCGATCGGCAATGACGGTCTG 1299

RESULT 7  
US-09-823-373-16  
; Sequence 16, Application US/09823373  
; Patent No. 6870038  
; GENERAL INFORMATION:  
; APPLICANT: Chaubhan, Sarita  
; APPLICANT: Diccolino, Robert  
; APPLICANT: Payne, Mark  
; APPLICANT: Gavagan, John  
; APPLICANT: Fallon, Robert  
; TITLE OF INVENTION: Isolation and Expression of a Gene for Nitrilase from  
; FILE REFERENCE: BC-1032 US NA  
; CURRENT APPLICATION NUMBER: US/09/823,373  
; CURRENT FILING DATE: 2001-03-30  
; PRIOR APPLICATION NUMBER: 60/193,707  
; NUMBER OF SEQ ID NOS: 32  
; SOFTWARE: Microsoft Office 97  
; SEQ ID NO 16  
; LENGTH: 1110  
; TYPE: DNA  
; ORGANISM: Artificial Sequence  
; FEATURE:  
; OTHER INFORMATION: Description of Artificial Sequence: A synthetic  
; OTHER INFORMATION: version of the nitrilase gene  
US-09-823-373-16

Query Match 21.4%; Score 216.8; DB 3; Length 1110;  
Best Local Similarity 51.9%; Pred. No. 1.1e-60;  
Matches 488; Conservative 0; Mismatches 452; Indels 0; Gaps 0;

14 TCAAGTCGCTGCGTCGCAAGCCGCCGATCTACATGATTGGAGCGACGGTGAACA 73  
23 TCTTGGCGGTACTGTTCAAGCTGAGCAGTTGGTTGGACGAGAGCGCTACATTAATGACA 82  
74 AAACCATTTGAGTGTATGGAAGAAAGACAGCATTAATTAATGCTGATGCTTTCCGG 133  
83 AGCTATCGTATCATGAAAGAAAGCTGCCAAAAGGGTGCTCTTTGATGCTTTCCAG 142  
134 AAACCTTGATTCAGAGCTACCCATGTTCTTTGGCTTGAATCCAGCATGGCAATGC 193  
143 AAGTTTATTCAGAGTTACCATATGCGCTGGTTGGTGAAGTTAAGTACTTTTGT 202  
194 AATTGTACGCGCAATACATGAGAACTGATGAGTTGATGGCTTCAAGCTAAGCGCA 253  
203 CCTTACTTCCAGATATACAGAACTTTGAGATTGGGTGACAGACAGAAATGGTGAAC 262  
254 TTTGAGATGCAAGCGGTTGGGATCATGTTCACTTGGGAGATGATGAACGGGTGC 313  
263 TGCAATTTGGCTGCCGTTGAAAACAAAATTTGCTTTGGTATGATGGGTTATTCGAGAGAG 322  
314 GTGGACCCCTTTACATCACTGCTGATGATGAGGATTAATGATGACCAATGGGGCCC 373  
323 CTGATATCTGTTACTGTTGCCAAGTCTTATCAGAGAGAGTGAATGTTGCAATC 382  
374 GGGAAAGTTGAAACCTACTTTTGTGAAGTACTTTGTTGCGGAGAGGGAGTGTTCAT 433  
383 GTGTAAGTTGAAGCCACTCAAGTTGAGCTACATCTACGAGAAAGTTAACGAAACCG 442  
434 CGTACCGGTTTTCAGACGCTGTTTGAAGGCTGGGAGCTTATGCTTTGGAGACAC 493  
443 ATTTCTTGACTCAGCAGCTTGCCTTCGAAAGTTGGTGAATTTGAATGTTGGAAACAT 502  
494 TTCAACCGGCTAACAAATAGCGCTTTGTATGACAAATATGAGATTCATTTGCGGCTT 553  
503 TCCAACTCTGTATTAAGTTATGATGATGATCTCTTGGGTAGAGCAAGTCCAGTTGCTT 562  
554 GGGCGAGCTTTAGCTTTATCTTAATGCGGCAAAAGCCCTGGGGCTGATGTCANTGAG 613  
563 GGCAGGCTATGTCCTCTTCAAGCCAGATGTTTTCATTTGTCATGAGCCAGACGCA 622  
614 CGGCTCTGATCTATGCGGTTGAAGGCAATGCTTCTGATGAGCTGTGCGCTG 673  
623 CCGTACCAATCTTACGCAATGGAAGGTCAAACTTTGCTTCTTCAACCAAGTCA 682  
674 TTTCACATTCATGATGATGATGTTGTACATGACGAAAGCAATGGTGGCTTCTG 733  
683 TTGGAACCTTCTGATGAAACCTTCTGTGAAAGCAAGAGAGCTTTGTTCCAC 742  
734 CTGCTGATGACATCACTGATCAATAGGCGCTGATGTTGTTGCTTGTGCGCCCTT 793  
743 AAGATGTTGGTGGGCAAGAAATTTACGTTCAAGATGATGAGCTTGGCCACTTT 802  
794 CCGAAATGAAAGAGGATTTCTCTAGCAAACTTTGATCTGAGTACGATCTTTGCTA 853  
803 CTGAAGATGCTGAAGGATTTTGTACCTGATGATGATTTGAGCAAAATTTCTGCGCA 862  
854 AAATGGCGGACAGCCCTGCTGATTAATCCGCTCCGACATTAATCTGCTGTATAG 913  
863 AGGCTGAGCGGATCACTGCTGCTGATCTTCAAGCTGAGCTTGTCCGTCCAGTTG 922  
914 ATCGAGCCCTAAATTAACCGGTAGTTGAATTTGAAGTTGA 953  
923 ACCCTAAGAACCACTCCAGTTCAAGAAATTTGATTTGA 962

## RESULT 8

US-08-447-702-4

; Sequence 4, Application US/08447702

; Patent No. 5629190

; GENERAL INFORMATION:

; APPLICANT: Petre, Dominique

; APPLICANT: Cerbelaud, Edith

APPLICANT: Levy-Schli, Sophie  
APPLICANT: Crouzet, Joel  
TITLE OF INVENTION: Polypeptides Possessing A Nitrlase  
TITLE OF INVENTION: Polypeptides Possessing A Nitrlase  
Patent No. 5629190  
TITLE OF INVENTION: Expression Cassettes and Host Microorganisms Enabling Them  
TITLE OF INVENTION: to be Obtained, and Method of Converting Nitrlases to Carboxy-  
NUMBER OF SEQUENCES: 5  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BURNS, DOANE, SWECKER & MATHEIS  
STREET: P. O. Box 1404  
CITY: Alexandria  
STATE: VA  
COUNTRY: USA  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/447,702  
FILING DATE: 23-MAY-1995  
CLASSIFICATION: 530  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: US 08/194,588  
FILING DATE: 10-FEB-1994  
PRIORITY APPLICATION DATA:  
APPLICATION NUMBER: FR 92-09-882  
FILING DATE: 10-AUG-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: McGowan, Malcolm K.  
REGISTRATION NUMBER: 39,300  
REFERENCE/DOCKET NUMBER: 003025-019  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 703-836-6620  
TELEFAX: 703-836-2021  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1194 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: unknown  
TOPOLOGY: unknown  
MOLECULE TYPE: DNA (genomic)  
FEATURE:  
NAME/KEY: CDS  
LOCATION: 87..1148  
US-08-447-702-4  
Query Match 13.1%; Score 132.4; DB 2; Length 1194;  
Best Local Similarity 48.1%; Pred. No. 7.6e-33;  
Matches 459; Conservative 0; Mismatches 471; Indels 24; Gaps 2;

14 TCAAGTCGCTGCGTCGCAAGCCGCCGATCTACATGATTGGAGCGACGGTGAACA 73  
106 TCAAGTGAAGCAAGTCAAGCTCTCTGATTTGAATCTGAGGCAACATGATATA 165  
74 AAACCATTTGAGTGTATGGAAGAAAGACAGCATTAATTAATGCTGATGCTTTCCGG 133  
166 AAACCTTGATTCAGAGCTACCCATGTTCTTTGGCTTGAATCCAGCATGGCAATGC 193  
134 AAACCTTGATTCAGAGCTACCCATGTTCTTTGGCTTGAATCCAGCATGGCAATGC 193  
226 AAGCATTTATTTCCGGCTATCCATTTGATTTGAGATCAAAATATGACTTCACTGAA 285  
194 AATTGTACGCGCAATACATGAGAACTGATGAGTTGATGGCCCTCAAGCTAAGCGCA 253  
286 TGATGTGGCGCTCTCTTTTCAAGAAATGCAATGAAATCCCAAGAAAGAAATTTCAACAAA 345  
254 TTTGAGATGCAAGCGGTTGGGATCATGTTCACTTGGGAGATGATGAACGGGTGC 313  
346 TTGATGATCTGCAAAAGAAATGAGTTTACGTTTGGGTTCTGTATACAGAAAGATA 405

314 GTGGACCCCTTACATGATGAGTTCATAGGCGATATAGTGACACATGGGGCC 373  
406 ATGCTCGCTATATTGACGCAATTTGTTGACCCGAAATGTAATTGATTGGCAAGC 465  
374 GAGCAAGTTGAAACCTATCTTTTGTGAACGTAATTTGTCGGCAGAGGGATGTTGAT 433  
466 ACAGGAATTTAAGCCACTAGTAGTAAGAGCTGTATGGGGAATGGGGATGGAACGA 525  
434 CGCTAGCGGTTTTCAGACGTCGTGTGGAAGGCTGGGTGCTTATGCTGTTGGAGCACC 493  
526 TGGCTCCGCTATTTTAAACAGAGTATGGGAATCTTGGGGGACTCCAGTCTGGGAACATG 585  
494 TTCAACCGCTAACAAATACGCTTTGTATGACAAATATGAAGATGATGTTGGGGCTT 553  
586 CTCTCCATTTAAACATTTGGGGGATGGCTCATTTGAACGAACGATACATGTTGCTTCT 645  
554 GAGCGAGCTTTAGCCTTTATCTTAAATGCGCGAAAGCCCTGGGGCTGATGTCATATGAG 613  
646 GGGCAGCTTGGCTTAAAGGCGAGATATCATCAGAGTATCATCAGCGCTGTGGGT 705  
614 CGGCTCTGCAATCTATGCGCTTGAAGGCAATGCTTC-----G 652  
706 CTACTAATGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 765  
653 TACTAGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 712  
766 TAAATATATCAACCAATCTGTTGGCCAGACATGATGATGATGATGATGATGATGATGAT 825  
713 AAAAGATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 772  
826 TTTCCAAAACCTTTCTACCGCTTGTCTGTAAGACACAGCATTTAT--TTCTAACACCG 882  
773 GTGACTTGTGCGCGCTTGTGCGGAAATGAAGAGTATTTCTTACGCAAACTTTGATC 832  
883 GTGAGATTTTGGCATCAATTCACACAGACGCGAGGATTTGCTGTGACAGATGATGAC 942  
833 CTGAGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 892  
943 TTAACCAATATATATATGAAAGTGTATGATGATGATGATGATGATGATGATGATGATG 1002  
893 ACATTACTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 946  
1003 GCTTCTTAAGTTTGAATTTGATGATGATGATGATGATGATGATGATGATGATGATG 1056

## RESULT 9

US-08-465-615-4

Patent No. 5635391  
Sequence 4, Application US/08465615

## GENERAL INFORMATION:

APPLICANT: PETRE, Dominique  
APPLICANT: CERBELAUD, Edith  
APPLICANT: LEVY-SCHILL, Sophie  
APPLICANT: CROUZET, JoelTITLE OF INVENTION: POLYPEPTIDES POSSESSING A NITRILASE  
TITLE OF INVENTION: ACTIVITY, DNA SEQUENCE CODING FOR SAID POLYPEPTIDES,  
TITLE OF INVENTION: EXPRESSION CASSETTES AND HOST MICROORGANISMS ENABLING THEM  
TITLE OF INVENTION: TO BE OBTAINED, AND METHOD OF CONVERTING NITRILES TO  
TITLE OF INVENTION: CARBOXYLATES BY MEANS OF SAID POLYPEPTIDE

## NUMBER OF SEQUENCES: 5

## CORRESPONDENCE ADDRESS:

ADDRESSEE: Burns, Doane, Swecker &amp; Mathis

STREET: P. O. Box 1404

CITY: Alexandria

STATE: Virginia

COUNTRY: United States

ZIP: 22313-1404

## COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: PatentIn Release #1.0, Version #1.25

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/465, 615  
FILING DATE:  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/194,588  
FILING DATE: 10-FEB-1994  
APPLICATION NUMBER: FR 9209882  
FILING DATE: 10-AUG-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Crane-Feury, Sharon E  
REGISTRATION NUMBER: 36,113  
REFERENCE/DOCKET NUMBER: 003025-015  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (703) 836-6620  
TELEFAX: (703) 836-2021  
INFORMATION FOR SEQ ID NO: 4:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 1194 base pairs  
TYPE: nucleic acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: DNA (genomic)  
US-08-465-615-4

Query Match 13.1%; Score 132.4; DB 2; Length 1194;  
Best Local Similarity 48.1%; Pred. No. 7.6e-33;  
Matches 459; Conservative 0; Mismatches 471; Indels 24; Gaps 2;  
14 TCAAGTGCCTGCGTCAAGCCGCCGATCTACATGATTTGAGGCGACGCTGAC 73  
106 TCAAGTGCAGAGCTGCAAGCTGCTCTGATTTATGAATCTAGGCAACAGTAGTA 165  
74 AAACATTTAGTATGAGAAAGACACGATTAATTAATGCTGCTGATGCTCTTCCG 133  
166 AAATTTAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 225  
134 AAATTTAGTATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 193  
226 AAGATTTATTTCCGGCTATCTCATATTTGATTTGAAATGATGATGATGATGATG 285  
194 AATTTAGCGCAATPACATGAGAACTATTGAGTTGATGAGCTTCAAGTAAAGCGCA 253  
286 TGAATGCGCGCTCTTTTCAAGATGCGATGGAATCCCAAGAAAGATTTCAACAA 345  
254 TTTGATGACAGCCAGCGGTTGGAAATCATGTCACCTGGGAGTATGAAAGCGCTCG 313  
346 TTAGTATGCTGCAAAAAGAAATGAGTTTACGTTTCTGTTCTGATCAGAAAGATA 405  
314 GTGCAACCTTTATCATGATGATGATGATGATGATGATGATGATGATGATGATGATG 373  
406 ATGCTCGCTATATTTGACGCAATTTGCTTGAACCCGAAATTTGATTTGAGCAAGC 465  
374 GCGAAAGTTGAACCTATTTTGTGAACGTAATTTGTCGGGAGGAGGATGTTGAT 433  
466 ACAGGAATTTAAGCCACTAGTAGTAAGAGCTGTATGGGGAATGGGGATGGAACGA 525  
434 CGCTAGCGGTTTTCAGACGTCGTGTGGAAGGCTGGGTGCTTATGCTGTTGGAGCACC 493  
526 TGGCTCCGCTATTTAAACAGAGTATGGGAATCTTGGGGGACTCCAGTCTGGGAACATG 585  
494 TTCAACCGCTAACAAATACGCTTTGTATGACAAATATGAAGATGATGTTGGGGCTT 553  
586 CTCTCCATTTAAACATTTGGGGGATGGCTCATTTGAACGAACGATACATGTTGCTTCT 645  
554 GAGCGAGCTTTAGCCTTTATCTTAAATGCGCGAAAGCCCTGGGGCTGATGTCATATGAG 613  
646 GGGCAGCTTGGCTTAAAGGCGAGATATCATCAGAGTATCATCAGCGCTGTGGGT 705  
614 CGGCTCTGCAATCTATGCGCTTGAAGGCAATGCTTC-----G 652  
706 CTACTAATGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 765





QY 259 GATCAGCAGCAAGCCGTTGGGAAATATAGTACCCCTGGAGATGATGAACGAGTCGGAGG 318  
Db 12182 GATGAGGAGGAGCCCTTTGGGAGTTAGCAACTTTGGGTATGATTAAGTACGTCCTTGGGAGC 1224  
QY 319 ACCCTTACATCAATCAATGATGTTTATAGGCGATATAGTGAACACATTTGGGCGCC 373  
Db 12242 AGGTGATCTTAAAGCAATGGTCTTCAAGTGTGATCTGTGACCCCTGGGGGGCTC 12296

```

RESULT 13
US-09-823-373-30/c
; Sequence 30, Application US/09823373
; Patent No. 6870038
; GENERAL INFORMATION:
; APPLICANT: Chauhan, Sarita
; APPLICANT: DiCosimo, Robert
; APPLICANT: Payne, Mark
; APPLICANT: Gavaagan, John
; APPLICANT: Fallon, Robert
; TITLE OF INVENTION: Isolation and Expression of a Gene for Nitritease from
; TITLE OF INVENTION: Acidovorax Facilis 72M
; FILE REFERENCE: BC-1032 US NA
; CURRENT APPLICATION NUMBER: US/09/823,373
; CURRENT FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 60/193,707
; PRIOR FILING DATE: 2000-03-31
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 30
; LENGTH: 90
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; IS-09-823-373-30

```

```

RESULT 14
US-09-823-373-27/c
; Sequence 27, Application US/09823373
; Patent No. 6870038
; GENERAL INFORMATION:
; APPLICANT: Chaunan, Sarita
; APPLICANT: DiCosimo, Robert
; APPLICANT: Payne, Mark
; APPLICANT: Gavagan, John
; APPLICANT: Fallon, Robert
; TITLE OF INVENTION: Isolation and Expression of a Gene for Nitrilase from
; TITLE OF INVENTION: Acidovorax Facitilis 72W
; FILE REFERENCE: BC-1032 US NA
; CURRENT APPLICATION NUMBER: US/09/823,373
; CURRENT FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 60/193,707
; PRIOR FILING DATE: 2000-03-31
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: Microsoft Office 97
; SEQ. ID NO. 27
; LENGTH: 90
; TYPE: DNA
; ORGANISM: Artificial Sequence

```

```

; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:Synthetic
; OTHER INFORMATION: oligonucleotide
US-09-823-373-27

```

Query Match	3.5%	Score 35;	DB 3;	length 90;
Best Local Similarity	63.9%	Pred. No.	0.23;	
Matches	53;	Conservative	0;	Mismatches
			30;	Indels
			0;	Gaps
			0;	

```

RESULT 15
US-09-823-373-32/c
; Sequence 32, Application US/09823373
; Patent No. 6870038
; GENERAL INFORMATION:
; APPLICANT: Chauban, Sarita
; APPLICANT: Dicosimo, Robert
; APPLICANT: Payne, Mark
; APPLICANT: Gavagan, John
; APPLICANT: Fallon, Robert
; TITLE OF INVENTION: Isolation and Expression of a Gene for Nitrlase from
; TITLE OF INVENTION: Acidovorax Facilis 72W
; FILE REFERENCE: BC-1032 US NA
; CURRENT APPLICATION NUMBER: US/09/823,373
; CURRENT FILING DATE: 2001-03-30
; PRIOR APPLICATION NUMBER: 60/193,707
; PRIOR FILING DATE: 2000-03-31
; NUMBER OF SEQ ID NOS: 32
; SOFTWARE: Microsoft Office 97
; SEQ ID NO 32
; LENGTH: 90
; TYPE: DNA
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: Description of Artificial Sequence:Synthetic
; IS-09-823-373-32

```

Search completed: April 27, 2006, 01:43:02  
Job time : 212.215 secs

**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 26, 2006, 23:12:44 ; Search time 976.006 Seconds

(without alignments)  
8591.290 Million cell updates/sec

Title: US-09-751-299-3

Perfect score: 1014

Sequence: 1 atgaagaagctatcaagc.....cgcaactcgaagaatttga 1014

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 9793542 seqs, 4134689005 residues

Total number of hits satisfying chosen parameters: 19587084

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database : Published Applications NA Main:\*

- 1: /cgn2\_6/ptodata/1/pubpna/US07\_PUBCOMB.seq.\*
- 2: /cgn2\_6/ptodata/1/pubpna/US08\_PUBCOMB.seq.\*
- 3: /cgn2\_6/ptodata/1/pubpna/US09A\_PUBCOMB.seq.\*
- 4: /cgn2\_6/ptodata/1/pubpna/US09B\_PUBCOMB.seq.\*
- 5: /cgn2\_6/ptodata/1/pubpna/US10A\_PUBCOMB.seq.\*
- 6: /cgn2\_6/ptodata/1/pubpna/US10B\_PUBCOMB.seq.\*
- 7: /cgn2\_6/ptodata/1/pubpna/US10C\_PUBCOMB.seq.\*
- 8: /cgn2\_6/ptodata/1/pubpna/US10D\_PUBCOMB.seq.\*
- 9: /cgn2\_6/ptodata/1/pubpna/US10E\_PUBCOMB.seq.\*
- 10: /cgn2\_6/ptodata/1/pubpna/US11\_PUBCOMB.seq.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1014	100.0	1014	3	US-09-751-299-3
2	1012.4	99.8	1014	6	US-10-146-772-385
3	1012.4	99.8	1014	6	US-10-241-742-385
4	1012.4	99.8	1014	7	US-10-440-523-385
5	1012.4	99.8	1014	7	US-10-440-523-385
6	1012.4	99.8	1014	7	US-10-440-523-385
7	1012.4	99.8	1014	7	US-10-461-925-385
8	1012.4	99.8	1014	7	US-10-146-772-47
9	1012.4	99.8	1014	6	US-10-241-742-47
10	1012.4	99.8	1014	7	US-10-440-523-47
11	1012.4	99.8	1014	7	US-10-440-523-47
12	1012.4	99.8	1014	6	US-10-146-772-57
13	1012.4	99.8	1014	6	US-10-241-742-57
14	1012.4	99.8	1014	6	US-10-440-523-57
15	1012.4	99.8	1014	7	US-10-440-523-57
16	1012.4	99.8	1014	7	US-10-461-925-57
17	1012.4	99.8	1014	6	US-10-146-772-139
18	1012.4	99.8	1014	6	US-10-241-742-139
19	1012.4	99.8	1014	7	US-10-440-523-139
20	1012.4	99.8	1014	7	US-10-440-523-139
21	1012.4	99.8	1014	7	US-10-461-925-139
22	1012.4	99.8	1014	6	US-10-146-772-291
23	1012.4	99.8	1014	6	US-10-241-742-291

24	396.4	39.1	1002	7	US-10-440-523-291	Sequence 291, App
25	396.4	39.1	1002	7	US-10-440-503-291	Sequence 291, App
26	396.4	39.1	1002	7	US-10-461-925-291	Sequence 291, App
27	364.2	35.9	1038	6	US-10-146-772-333	Sequence 333, App
28	364.2	35.9	1038	6	US-10-241-742-333	Sequence 333, App
29	364.2	35.9	1038	7	US-10-440-523-333	Sequence 333, App
30	364.2	35.9	1038	7	US-10-440-503-333	Sequence 333, App
31	364.2	35.9	1038	7	US-10-461-925-333	Sequence 333, App
32	358.6	35.4	1068	6	US-10-146-772-305	Sequence 305, App
33	358.6	35.4	1068	6	US-10-241-742-305	Sequence 305, App
34	358.6	35.4	1068	7	US-10-440-523-305	Sequence 305, App
35	358.6	35.4	1068	7	US-10-440-503-305	Sequence 305, App
36	358.6	35.4	1068	7	US-10-461-925-305	Sequence 305, App
37	319	31.5	1011	6	US-10-146-772-217	Sequence 217, App
38	319	31.5	1011	6	US-10-241-742-217	Sequence 217, App
39	319	31.5	1011	7	US-10-440-523-217	Sequence 217, App
40	319	31.5	1011	7	US-10-440-503-217	Sequence 217, App
41	319	31.5	1011	7	US-10-461-925-217	Sequence 217, App
42	311.4	30.7	1026	6	US-10-146-772-33	Sequence 33, App1
43	311.4	30.7	1026	6	US-10-241-742-33	Sequence 33, App1
44	311.4	30.7	1026	6	US-10-440-523-33	Sequence 33, App1
45	311.4	30.7	1026	7	US-10-440-503-33	Sequence 33, App1

## ALIGNMENTS

RESULT 1  
US-09-751-299-3  
Sequence 3, Application US/09751299  
Patent No. US20020012974A1  
GENERAL INFORMATION:  
APPLICANT: Madden, Mark  
APPLICANT: Weiner, David P.  
TITLE OF INVENTION: METHODS FOR PRODUCING ENANTIOMERICALLY PURE  
TITLE OF INVENTION: ALPHA-SUBSTITUTED CARBOXYLIC ACIDS  
FILE REFERENCE: DIVER1440-2  
CURRENT APPLICATION NUMBER: US/09/751,299  
CURRENT FILING DATE: 2000-12-29  
PRIOR APPLICATION NUMBER: 60/254,414  
PRIOR FILING DATE: 2000-12-07  
PRIOR APPLICATION NUMBER: 60/173,609  
PRIOR FILING DATE: 1999-12-29  
NUMBER OF SEQ ID NOS: 4  
SOFTWARE: PatentIn Ver. 2.1  
SEQ ID NO 3  
LENGTH: 1014  
TYPE: DNA  
ORGANISM: Unknown Organism  
FEATURE:  
OTHER INFORMATION: Description of Unknown Organism: Obtained from an  
OTHER INFORMATION: environmental sample  
NAME/KEY: CDS  
LOCATION: (1)..(1014)  
US-09-751-299-3  
Query Match 100.0%; Score 1014; DB 3; Length 1014;  
Best Local Similarity 100.0%; Pred. No. 0;  
Matches 1014; Conservative 0; Mismatches 0; Indels 0; Gaps 0;  
QY 1 ATGAAAGCTATCAAGTCCTGCTGGAACCCCGCATCTACATGATTTGGAG 60  
DB 1 ATGAAAGCTATCAAGTCCTGCTGGAACCCCGCATCTACATGATTTGGAG 60  
QY 61 GCGACGGTGACAAACATTGATGATGAGAGACAGACCTAATATGCTGCTG 120  
DB 61 GCGACGGTGACAAACATTGATGATGAGAGACAGACCTAATATGCTGCTG 120  
QY 121 ATGCGCTTTCCGGAACCTGATTCAGGCTACCATGATTTCTTGGCTTACTACCA 180  
DB 121 ATGCGCTTTCCGGAACCTGATTCAGGCTACCATGATTTCTTGGCTTACTACCA 180





```
APPLICANT: Milan, Aileen
APPLICANT: Desantis, Grace
APPLICANT: Madden, Mark
APPLICANT: Burk, Mark
TITLE OF INVENTION: Nitrites
FILE REFERENCE: Docket No. US20040014195A1 DIV-013US
CURRENT APPLICATION NUMBER: US/10/440.523
PRIOR FILING DATE: 2003-05-15
PRIOR APPLICATION NUMBER: US/10/146.772
PRIOR FILING DATE: 2002-05-15
PRIOR APPLICATION NUMBER: US 60/309,006
PRIOR FILING DATE: 2001-07-30
PRIOR APPLICATION NUMBER: US 60/351,336
PRIOR FILING DATE: 2002-01-22
PRIOR APPLICATION NUMBER: US 60/300,189
PRIOR FILING DATE: 2001-06-21
PRIOR APPLICATION NUMBER: US 09/751,299
PRIOR FILING DATE: 2000-12-28
PRIOR APPLICATION NUMBER: US 60/254,414
PRIOR FILING DATE: 2000-12-07
PRIOR APPLICATION NUMBER: US 60/173,609
PRIOR FILING DATE: 1999-12-29
NUMBER OF SEQ ID NOS: 386
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 385
LENGTH: 1014
TYPE: DNA
ORGANISM: Unknown
FEATURE:
OTHER INFORMATION: Obtained from an environmental sample
US-10-440-523-385
```

```
Query Match 99.8%; Score 1012.4; DB 7; Length 1014;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1013; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGAAAGAGCTATCAAGTGCCTGCTGCAAGCCGCCGATCATGATGATTTGAG 60
DB 1 ATGAAAGAGCTATCAAGTGCCTGCTGCAAGCCGCCGATCATGATGATTTGAG 60
QY 61 GCGACGGTGCACAAACCATTTGATGATGAAAGAGCAGCAGTAATATGCTGCTG 120
DB 61 GCGACGGTGCACAAACCATTTGATGATGAAAGAGCAGCAGTAATATGCTGCTG 120
QY 121 ATGCGCTTCCGGAATCTTGATTCAGGCTTCCATGATTTCTTTGGCTGACTACCA 180
DB 121 ATGCGCTTCCGGAATCTTGATTCAGGCTTCCATGATTTCTTTGGCTGACTACCA 180
QY 181 GCATGGGCAATGCAATTTGTATGAGCCATATCCATGAACTCATTTGAGTGGATGCCCT 240
DB 181 GCATGGGCAATGCAATTTGTATGAGCCATATCCATGAACTCATTTGAGTGGATGCCCT 240
QY 241 CAAAGTAAAGCGATTTGATGATGACCAAGCGGTTGGAAATCATGCTACCTGGGAGATG 300
DB 241 CAAAGTAAAGCGATTTGATGATGACCAAGCGGTTGGAAATCATGCTACCTGGGAGATG 300
QY 301 AGTGAACGGGTGCTGAGCACTTTACATCACTGATGATGATGAGGATATGATGAC 360
DB 301 AGTGAACGGGTGCTGAGCACTTTACATCACTGATGATGATGAGGATATGATGAC 360
QY 361 ACCATTTGGGGCCCGCGAAAGTTGAACCTTATTTGTAACGTAATTTGTTCCGCCGA 420
DB 361 ACCATTTGGGGCCCGCGAAAGTTGAACCTTATTTGTAACGTAATTTGTTCCGCCGA 420
QY 421 GGGGATGCTATCGCTAGCGGTTTTCAGAGAGTCTGTTGGAAGGCTGGGTGCTTATGC 480
DB 421 GGGGATGCTATCGCTAGCGGTTTTCAGAGAGTCTGTTGGAAGGCTGGGTGCTTATGC 480
QY 481 TGTGGGAGCAGCTTCAACCGCTTAAACAAATACGCTTTGTATGACAAATATGAGATTT 540
DB 481 TGTGGGAGCAGCTTCAACCGCTTAAACAAATACGCTTTGTATGACAAATATGAGATTT 540
QY 541 CATTTGCGGCTTGGCCGAGCTTTAGCTTTATCTTAATGCGCGAAAGCCTCGGGGCT 600
```

```
|||||
DB 541 CATTTGCGGCTTGGCCGAGCTTTAGCTTTATCTTAATGCGCGAAAGCCTCGGGGCT 600
QY 601 GATGTCATATGACGGGCTTCCGAATCTATGCGGTTGAAGGGCAATGCTTGTACTAGC 660
DB 601 GATGTCATATGACGGGCTTCCGAATCTATGCGGTTGAAGGGCAATGCTTGTACTAGC 660
QY 661 TCGTGTGCTGCTGTTTCAACATCATGATGATGATGCTTTGTACAGATGAGAAACAT 720
DB 661 TCGTGTGCTGCTGTTTCAACATCATGATGATGATGCTTTGTACAGATGAGAAACAT 720
QY 721 GCGTTGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 780
DB 721 GCGTTGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 780
QY 781 GTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 840
DB 781 GTGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 840
QY 841 CGCATCTTGTCTAAATATGAGGCGGAGACCTGCTGCTGCTGCTGCTGCTGCTGCTG 900
DB 841 CGCATCTTGTCTAAATATGAGGCGGAGACCTGCTGCTGCTGCTGCTGCTGCTGCTG 900
QY 901 CGCTTGTCTAATAGATCCGAGCCCTAAATTAACCGGTATGTAATGAAAGGATCTTCT 960
DB 901 CGCTTGTCTAATAGATCCGAGCCCTAAATTAACCGGTATGTAATGAAAGGATCTTCT 960
QY 961 CCTTACGCTTTGGGTAAAGGCTGAGAGCGGTCGCAACCTCGAAGAAATTTGA 1014
DB 961 CCTTACGCTTTGGGTAAAGGCTGAGAGCGGTCGCAACCTCGAAGAAATTTGA 1014
```

```
RESULT 5
US-10-440-503-385
Sequence 385, Application US/10440503
Publication No. US20040038419A1
GENERAL INFORMATION:
APPLICANT: Weiner, David Paul
APPLICANT: Chaplin, Jennifer Ann
APPLICANT: Chi, Ellen
APPLICANT: Milan, Aileen
APPLICANT: Desantis, Grace
APPLICANT: Burk, Mark J.
APPLICANT: McQuaid, Jeffrey
APPLICANT: Stege, Justin
TITLE OF INVENTION: ASSAYS AND KITS FOR DETECTING THE PRESENCE OF
FILE REFERENCE: 09010-900001
CURRENT APPLICATION NUMBER: US/10/440.503
PRIOR FILING DATE: 2003-05-15
PRIOR APPLICATION NUMBER: US 60/380,737
PRIOR FILING DATE: 2002-05-15
NUMBER OF SEQ ID NOS: 386
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 385
LENGTH: 1014
TYPE: DNA
ORGANISM: Unknown
FEATURE:
OTHER INFORMATION: Obtained from an environmental sample
US-10-440-503-385
```

```
Query Match 99.8%; Score 1012.4; DB 7; Length 1014;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1013; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGAAAGAGCTATCAAGTGCCTGCTGCAAGCCGCCGATCATGATGATTTGAG 60
DB 1 ATGAAAGAGCTATCAAGTGCCTGCTGCAAGCCGCCGATCATGATGATTTGAG 60
QY 61 GCGACGGTGCACAAACCATTTGATGATGAAAGAGCAGCAGTAATATGCTGCTG 120
DB 61 GCGACGGTGCACAAACCATTTGATGATGAAAGAGCAGCAGTAATATGCTGCTG 120
```

```
QY 121 ATGCGCTTCCGGAAACTTGGATTCCAGGCTACCATGGTTCTTTGGCTGACTCACA 180
Db 121 ATGCGCTTCCGGAAACTTGGATTCCAGGCTACCATGGTTCTTTGGCTGACTCACA 180
QY 181 GCATGGGCAATCAATTTGTACGCCAATCCATGAGAACTCATTTGGATGGATGGCCCT 240
Db 181 GCATGGGCAATCAATTTGTACGCCAATCCATGAGAACTCATTTGGATGGATGGCCCT 240
QY 241 CAAGCTAAGCGCATTTGAGATGACGACGCAAGCGGTGGGAATCATGTGACCTCGGGATG 300
Db 241 CAAGCTAAGCGCATTTGAGATGACGACGCAAGCGGTGGGAATCATGTGACCTCGGGATG 300
QY 301 AGTGAACGGGTGGGTGGCAACCTTTACATCAGTCAAGTGTTCATAGGCAATATGTGAC 360
Db 301 AGTGAACGGGTGGGTGGCAACCTTTACATCAGTCAAGTGTTCATAGGCAATATGTGAC 360
QY 361 ACCATTGGGGGCCCGGGAAGTTGAACCTTACTTTTGTGAAAGTACTTTGTTGGCGCA 420
Db 361 ACCATTGGGGGCCCGGGAAGTTGAACCTTACTTTTGTGAAAGTACTTTGTTGGCGCA 420
QY 421 GGGGATGGTTCAATCGCTACCGGTTTTCAGACGTCGTGTTGAAAGCTGGGTGGCTTATGC 480
Db 421 GGGGATGGTTCAATCGCTACCGGTTTTCAGACGTCGTGTTGAAAGCTGGGTGGCTTATGC 480
QY 481 TGTGGGAGACCTTCAACCGCTTAACAAATACGCTTTGTATGCAAAATGAGAGATT 540
Db 481 TGTGGGAGACCTTCAACCGCTTAACAAATACGCTTTGTATGCAAAATGAGAGATT 540
QY 541 CATTTGGCGGCTTGGCCGAGCTTTAGCCTTTATCTTATGCGGCGAAGCCCTGGGGCCT 600
Db 541 CATTTGGCGGCTTGGCCGAGCTTTAGCCTTTATCTTATGCGGCGAAGCCCTGGGGCCT 600
QY 601 GATGTCATATGATAGCGGCTCTCGAATCTATGCGGTTGAAGGGCAATGCTTCTACTAGG 660
Db 601 GATGTCATATGATAGCGGCTCTCGAATCTATGCGGTTGAAGGGCAATGCTTCTACTAGG 660
QY 661 TGTGTGGCGCTGTTTCAATTCATGATGCAATGCTTTGTATGACATGCAAAAGCAT 720
Db 661 TGTGTGGCGCTGTTTCAATTCATGATGCAATGCTTTGTATGACATGCAAAAGCAT 720
QY 721 GGGTGGCTTGGCGGTGGTGAACCTGAGATATAGGGCTGATGGTGTGACTTG 780
Db 721 GGGTGGCTTGGCGGTGGTGAACCTGAGATATAGGGCTGATGGTGTGACTTG 780
QY 781 GTGCGGCTCTTGGCGAAATGAGAGGATATCTCTACGAAACCTTGATCTGTGAGTA 840
Db 781 GTGCGGCTCTTGGCGAAATGAGAGGATATCTCTACGAAACCTTGATCTGTGAGTA 840
QY 841 GCGATCTTGTGAATGGCGGACACCTGCTGTGATTTCCCGTCCGACATTAAT 900
Db 841 GCGATCTTGTGAATGGCGGACACCTGCTGTGATTTCCCGTCCGACATTAAT 900
QY 901 GCGTGTCTATATGATGCGAGCCCTTAATTAACGGTATGGAATGGAAGTATCTTGGT 960
Db 901 GCGTGTCTATATGATGCGAGCCCTTAATTAACGGTATGGAATGGAAGTATCTTGGT 960
QY 961 CCTTACGCTTGGGTGAAGCGTCTGAGACGGGTGGCGCACTGGAAGAAATTTGA 1014
Db 961 CCTTACGCTTGGGTGAAGCGTCTGAGACGGGTGGCGCACTGGAAGAAATTTGA 1014

RESULT 6
US-10-461-925-385
; Sequence 385, Application US/10461925
; Publication No. US20040053378A1
; GENERAL INFORMATION:
; APPLICANT: Mark J. Burk
; APPLICANT: Desantis, Grace
; APPLICANT: Morgan, Brian
; APPLICANT: Zhu, Zoulin
; TITLE OF INVENTION: PROCESSES FOR MAKING (R) - ETHYL 4-CYANO-3-HYDROXYBUTYRIC ACID
; FILE REFERENCE: 09010-270001
```

```
; CURRENT APPLICATION NUMBER: US/10/461,925
; CURRENT FILING DATE: 2003-06-13
; PRIOR APPLICATION NUMBER: US 60/389,317
; PRIOR FILING DATE: 2002-06-13
; PRIOR APPLICATION NUMBER: US 60/392,944
; PRIOR FILING DATE: 2002-06-28
; NUMBER OF SEQ ID NOS: 386
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 385
; LENGTH: 1014
; TYPE: DNA
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Obtained from an environmental sample
; US-10-461-925-385

Query Match          99.8%; Score 1012.4; DB 7; Length 1014;
Best Local Similarity 99.9%; Pred. No. 0;
Matches 1013; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 ATGAAAGAAGCTATCAAGTGGCTGCGTGCAGCGCCCGCATTTACATGAGATTGAG 60
Db 1 ATGAAAGAAGCTATCAAGTGGCTGCGTGCAGCGCCCGCATTTACATGAGATTGAG 60
QY 61 GCGAGGTTGGAACAAACCATTTGATGATGGAAGAGACAGCATATATGCTGCTTG 120
Db 61 GCGAGGTTGGAACAAACCATTTGATGATGGAAGAGACAGCATATATGCTGCTTG 120
QY 121 ATGCGCTTCCGGAAACTTGGATTCCAGGCTACCATGGTTCTTTGGCTGACTCACA 180
Db 121 ATGCGCTTCCGGAAACTTGGATTCCAGGCTACCATGGTTCTTTGGCTGACTCACA 180
QY 181 GCATGGGCAATCAATTTGTACGCCAATCCATGAGAACTCATTTGGATGGATGGCCCT 240
Db 181 GCATGGGCAATCAATTTGTACGCCAATCCATGAGAACTCATTTGGATGGATGGCCCT 240
QY 241 CAAGCTAAGCGCATTTGAGATGACGACGCAAGCGGTGGGAATCATGTGACCTCGGGATG 300
Db 241 CAAGCTAAGCGCATTTGAGATGACGACGCAAGCGGTGGGAATCATGTGACCTCGGGATG 300
QY 301 AGTGAACGGGTGGGTGGCAACCTTTACATCAGTCAAGTGTTCATAGGCAATATGTGAC 360
Db 301 AGTGAACGGGTGGGTGGCAACCTTTACATCAGTCAAGTGTTCATAGGCAATATGTGAC 360
QY 361 ACCATTGGGGGCCCGGGAAGTTGAACCTTACTTTTGTGAAAGTACTTTGTTGGCGCA 420
Db 361 ACCATTGGGGGCCCGGGAAGTTGAACCTTACTTTTGTGAAAGTACTTTGTTGGCGCA 420
QY 421 GGGGATGGTTCAATCGCTACCGGTTTTCAGACGTCGTGTTGAAAGCTGGGTGGCTTATGC 480
Db 421 GGGGATGGTTCAATCGCTACCGGTTTTCAGACGTCGTGTTGAAAGCTGGGTGGCTTATGC 480
QY 481 TGTGGGAGACCTTCAACCGCTTAACAAATACGCTTTGTATGCAAAATGAGAGATT 540
Db 481 TGTGGGAGACCTTCAACCGCTTAACAAATACGCTTTGTATGCAAAATGAGAGATT 540
QY 541 CATTTGGCGGCTTGGCCGAGCTTTAGCCTTTATCTTATGCGGCGAAGCCCTGGGGCCT 600
Db 541 CATTTGGCGGCTTGGCCGAGCTTTAGCCTTTATCTTATGCGGCGAAGCCCTGGGGCCT 600
QY 601 GATGTCATATGATAGCGGCTCTCGAATCTATGCGGTTGAAGGGCAATGCTTCTACTAGG 660
Db 601 GATGTCATATGATAGCGGCTCTCGAATCTATGCGGTTGAAGGGCAATGCTTCTACTAGG 660
QY 661 TGTGTGGCGCTGTTTCAATTCATGATGCAATGCTTTGTATGACATGCAAAAGCAT 720
Db 661 TGTGTGGCGCTGTTTCAATTCATGATGCAATGCTTTGTATGACATGCAAAAGCAT 720
QY 721 GGGTGGCTTGGCGGTGGTGAACCTGAGATATAGGGCTGATGGTGTGACTTG 780
Db 721 GGGTGGCTTGGCGGTGGTGAACCTGAGATATAGGGCTGATGGTGTGACTTG 780
QY 781 GTGCGGCTCTTGGCGAAATGAGAGGATATCTCTACGAAACCTTGATCTGTGAGTA 840
```

```

Db      781 GTCCGCGCTCTTGGCGAAATGAAGAGGTATCTCTACGCAACCTTGATCTGAGTA 840
Qy      841 CGCATCTTGTATAAATGGGGGAGACCTGTGTGATTAATCCGTCGCCGACATTACT 900
Db      841 CGCATCTTGTATAAATGGGGGAGACCTGTGTGATTAATCCGTCGCCGACATTACT 900
Qy      901 CGCTTGCTAATAGATCGAGCCCTTAATTTACCGGTAGTTGAATTTGAAGGTGATCTTCGT 960
Db      901 CGCTTGCTAATAGATCGAGCCCTTAATTTACCGGTAGTTGAATTTGAAGGTGATCTTCGT 960
Qy      961 CCTTACGTTTGGGTAAAGCGCTGAGAGCGGTGCGCAACTCGAAGAAATTTGA 1014
Db      961 CCTTACGTTTGGGTAAAGCGCTGAGAGCGGTGCGCAACTCGAAGAAATTTGA 1014

RESULT 7
US-10-146-772-47
; Sequence 47, Application US/10146772
; Publication No. US20030124698A1
; GENERAL INFORMATION:
; APPLICANT: Short, Jay
; APPLICANT: Weiner, David
; APPLICANT: Chaplin, Jennifer
; APPLICANT: Chi, Ellen
; APPLICANT: Milan, Aileen
; APPLICANT: Desantis, Grace
; APPLICANT: Madden, Mark
; APPLICANT: Burk, Mark
; TITLE OF INVENTION: Nitriases
; FILE REFERENCE: Docket No. US20030124698A1 DIV-013US
; CURRENT APPLICATION NUMBER: US/10/146,772
; CURRENT FILING DATE: 2002-05-15
; PRIOR APPLICATION NUMBER: US 60/309,006
; PRIOR FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US 60/351,336
; PRIOR FILING DATE: 2002-01-22
; PRIOR APPLICATION NUMBER: US 60/300,189
; PRIOR FILING DATE: 2001-06-21
; PRIOR APPLICATION NUMBER: US 09/751,299
; PRIOR FILING DATE: 2000-12-28
; PRIOR APPLICATION NUMBER: US 60/254,414
; PRIOR FILING DATE: 2000-12-07
; PRIOR APPLICATION NUMBER: US 60/173,609
; PRIOR FILING DATE: 1999-12-29
; NUMBER OF SEQ ID NOS: 386
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 47
; LENGTH: 1014
; TYPE: DNA
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Obtained from an environmental sample
US-10-146-772-47

Query Match      49.3%; Score 499.4; DB 6; Length 1014;
Best Local Similarity 68.3%; Pred. No. 1.6e-16;
Matches 692; Conservative 0; Mismatches 321; Indels 0; Gaps 0;

```

```

Db      182 CTTGAGGAGATGCAATTCGTGCGCCGATACCAAGAAATCTACGTGCTCTGACAGCCCTC 241
Qy      242 AAGCTAAGCGCATTTGAGATGACAGCCAGGCGTTGGAAATCATGTGTACCTTGGGATGA 301
Db      242 AAGCCAAAGCGCATGAGTGAAGGCGCCGACGCGCGGTATATACGTGCGCTGAGGATGA 301
Qy      302 GTGAACGGGTGCGGTGACACCTTTACATCAATGATGAGTGTATATAGGCGATATGTGACA 361
Db      302 GTGAACGGGTGAGGAAACCTTTACATGAGGCGAGTGGTCAATTGAGATTAAGGCGAAA 361
Qy      362 CCATTGGGGCGCGGCGAAAGTTGAACCTTCTTTGTTGAACGTATCTTTGTGGCGGAG 421
Db      362 CAGCTGGGCGTGGCGGAAAGCTGAACCAACCATGTAGAGCGAACCTTCTTGGTGAAG 421
Qy      422 GGAATGTTTATCGTATACGGTTTTCGAGAGTGTGTTGAAGGCTGGTGGCTTATGCT 481
Db      422 GCGACGGATCATCTCTTTCACATTTCGACACACCGTTGGGGGTCTGGGCGGACTGCT 481
Qy      482 GTTGGGAGACCTTCAACCGCTAACMAAATACGTTTGTATGACMAAATGAAGATTC 541
Db      482 GTTGGGAAACCTTACACCTTTTGAATATGCGCTCTACGACACAGAACGAGAAATAC 541
Qy      542 ATTGTCGGCTTGGCGGAGCTTTAGCCTTTATCCATATGCGGCGAAACCTTGGGCGCTG 601
Db      542 ACTTCGCGCGCTGGCGCTTACGATCTACGATCTACGATCAAGACAGAAATCTTGGACAG 601
Qy      602 ATGTCAATGATGAGCGGCTCTCGAATCTATGCGGTTGAAGGCGAATGCTTGTACTAGCT 661
Db      602 AAGTAAATGTGCAAGCTTCTCGGATCTAGCGGTGAAGGCGAGTGTGTTGTTCTCGCTT 661
Qy      662 CGTGTGGCTGTTTCACAATCATGATGATATGATGTTGTATGATGACGAATGACGAATG 721
Db      662 CTTGGCGCTGCTGTGCGGACGAGATGATGAAATGCTTGCATGACGAATGACGAATG 721
Qy      722 GCTTCTTTCAGGCGCGGCGGCGGTATCCCGCATTTATCGTCCGATGCGACGACCTAG 781
Db      722 GCTTCTTTCAGGCGCGGCGGCGGTATCCCGCATTTATCGTCCGATGCGACGACCTAG 781
Qy      782 TCGGCGCTTTCGCGAAATGAAGAGGTATTTCTTACCGCAACCTTGATCTGAGTAC 841
Db      782 CGCGCCCTTTCGCGGAAAGAGAGGTATTTCTTACCGCAACCTTGATCTGAGTAC 841
Qy      842 GCATCTTGTGTAATAGCGGCGGACCTGCTGTATATTCGCTCCGACATTACTC 901
Db      842 GAATCTTGTGTAATAGCGGCGGACCTGCTGTATATTCGCTCCGACATTACTC 901
Qy      902 GCTTGCTAATGATGAGCGGCTTAATTTACCGGTAGTTGAATTTGAAGGTGATCTTCGT 961
Db      902 GCGTGTGATCATGCGAGTGCCTAATCAGCCAGTGTGAGGTTGGAAGGAAATACCTG 961
Qy      962 CTTACGCTTGGGTAAAGCGCTGAGAGCGGTGCGCAACTCGAAGAAATTTGA 1014
Db      962 CATCGGCCAAAGGCTTTGAAGTTGAGGCGGCGCCCGGAGTACGAAGGCGATTTGA 1014

RESULT 8
US-10-241-742-47
; Sequence 47, Application US/10241742
; Publication No. US20040002147A1
; GENERAL INFORMATION:
; APPLICANT: Short, Jay
; APPLICANT: Weiner, David
; APPLICANT: Chaplin, Jennifer
; APPLICANT: Chi, Ellen
; APPLICANT: Milan, Aileen
; APPLICANT: Desantis, Grace
; APPLICANT: Madden, Mark
; APPLICANT: Burk, Mark
; TITLE OF INVENTION: Nitriases
; FILE REFERENCE: Docket No. US20040002147A1 DIV-013US
; CURRENT APPLICATION NUMBER: US/10/241,742
; CURRENT FILING DATE: 2002-09-09
; PRIOR APPLICATION NUMBER: US/10/146,772

```



Qy	2	TGAAAGAGCTATCAAGAGTCGCTGCGTGCAGAACCGGCCCATCTCAATGCAATTTGGAGG	61
Db	2	TGAAAGAGGAATCAAGAGTAGCGCTGTGTCAAGACAGCTCCAGTCTTTCTTCACCTCGAAG	61
Qy	62	CGACGGTGCACAAACCATTTAGATTATGAAAGAACGACAGATGAATTAATCTGTCGTA	121
Db	62	CCACAGTGCACAAAGCCGTGCGCCCTGATTGAGAGAGCAGCCGTATACGGCGCACGCTTA	121
Qy	122	TGCGCTTTCCGAAACTGTGATTCAGGCTACAGGCTACATGGTTCCTTTGCTTGACTCACAG	181
Db	122	TGCGCTTTCCAGAGACCTGGATTCCAGGCTACCCATGGTTCCTTTGCTTGACTCACAG	181
Qy	182	CATGGCGAAAGCAATTTGTATCCGCATATACATAGAAATCTCATTTGAGATTGAGATGCCCTC	241
Db	182	CTGGGGGAGTCAATTCGTGCGCCGATTCACAGAAATCTCACTGTCCTTCCACAGCCCTC	241
Qy	242	AAGCTAAGCGCATTTTCAGATGACAGCCAAACGGTTGGAGATCATGGTCACTCCGTGGAGATGA	301
Db	242	AGGCGAAGCGCATCAAGTAGAGGCCGCCAGCGCGCGATATACGTGCGCTTAGGGATACA	301
Qy	302	GTAACGGGTCGGTGGCACCTTTACATCAAGTCAGTGGATTCAATAGCGATATAGTACA	361
Db	302	GCGAAACGCGTGAACGGGAACCTCTTCATATGGGCGATGGCTCATTTGACATATAGGGCGAAA	361
Qy	362	CCATTTGGGGGCCCGGCGAAGTTGAAACCTACTTTTGTGAACGTACTTTGTTCCGCGAAG	421
Db	362	CAGCTGGGCTGCGCCCAAGAGCTGAACCAACCATGTATAGACGAAACCTCTTTCGGTAAAG	421
Qy	422	GGGATGGTTATCGCTAGCGGTTTTGGAACGCTCGTTGGAAGCGTGGTGGCTTATGCT	481
Db	422	GCGACGATATCTCTTTCACTTTGACACACCTTTGGGGGTGTGGGCGGAACTTGCT	481
Qy	482	GTTGGGAGCACTTCAACCGCTPAACAAATACGCTTTGTATGACAAATATGAAGATTC	541
Db	482	GTTGGGAACATTACAACCTCTTTGGAATATAGCGCTTACGACAGAACGAGGAATATAC	541
Qy	542	ATTGTGCGGTTGGCCGAGCTTTAGCTTTATCTTAATGGCGCGAAAGCCCTGGGCGCTG	601
Db	542	ACTTCGCGCTGGCTTACCTTCAGATCTACCGTCAAGCAAGAAATCTTTGGACAG	601
Qy	602	ATGTCAATGATGCGGCTCTGCATATCTATGCGTTGAAGGCAATGCTTCTGATTAAGCGT	661
Db	602	AAGTAATATGTGCAAGCTTCTCGATCTACGCCGTGGAAGGCAAGTTTGTCTCGCTT	661
Qy	662	CGTGTGGCTGTTTCAACATCAATCAATGATATGCTTTGTATCAAGATGACGAAAGCATG	721
Db	662	CGTGTGGCTGTTTCAACATCAATCAATGATATGCTTTGTATCAAGATGACGAAAGCATG	721

```

/ APPLICANT: Short, Jay
/ APPLICANT: Weiner, David
/ APPLICANT: Chaplin, Jennifer
/ APPLICANT: Chi, Ellen
/ APPLICANT: Milan, Aileen
/ APPLICANT: Desantis, Grace
/ APPLICANT: Madden, Mark
/ APPLICANT: Burk, Mark
/ TITLE OF INVENTION: Nitrlases
/ FILE REFERENCE: Docket No. US20040014195A1 DIV-013US
/ CURRENT APPLICATION NUMBER: US/10/440,523
/ CURRENT FILING DATE: 2003-05-15
/ PRIOR APPLICATION NUMBER: US/10/146,772
/ PRIOR FILING DATE: 2002-05-15
/ PRIOR APPLICATION NUMBER: US 60/309,006
/ PRIOR FILING DATE: 2001-07-30
/ PRIOR APPLICATION NUMBER: US 60/351,336
/ PRIOR FILING DATE: 2002-01-22
/ PRIOR APPLICATION NUMBER: US 60/300,189
/ PRIOR FILING DATE: 2001-06-21
/ PRIOR APPLICATION NUMBER: US 09/751,299
/ PRIOR FILING DATE: 2000-12-28
/ PRIOR APPLICATION NUMBER: US 60/254,414
/ PRIOR FILING DATE: 2000-12-07
/ PRIOR APPLICATION NUMBER: US 60/173,609
/ PRIOR FILING DATE: 1999-12-29
/ NUMBER OF SEQ ID NOS: 386
/ SOFTWARE: FastSeq for Windows Version 4.0
/ SEQ ID NO 47
/ LENGTH: 1014
/ TYPE: DNA
/ ORGANISM: Unknown
/ FEATURE:
/ OTHER INFORMATION: Obtained from an environmental sample
/ US-10-440-523-47

```

Query Match	49.3%	Score 499.4	DB 7	Length 1014
Best Local Similarity	68.3%	Pred. No. 1.6e-161		
Matches 692	Conservative 0	Mismatches 321	Indels 0	Gaps 0
OY	2	TGAAGAAGCTATTAAGTGGCTGCGTGCAGCGCCCGCATTTACATGATTTGAGG	61	
Db	2	TGAAGAAGACAAATTAAGTGGCTGCTGTGACAGCAGTCCAGTCTTTCGACCTGGACG	61	
OY	62	CGACGGTGGACAAAAACCATTTGATTTGATGGAAGAACGACAGCTAATTAATGCTGTGTGA	121	





PRIOR APPLICATION NUMBER: US 60/254,414  
 PRIOR FILING DATE: 2000-12-07  
 PRIOR APPLICATION NUMBER: US 60/173,609  
 PRIOR FILING DATE: 1999-12-29  
 NUMBER OF SEQ ID NOS: 386  
 SOFTWARE: FastSeq for Windows Version 4.0  
 SEQ ID NO 57  
 LENGTH: 1014  
 TYPE: DNA  
 ORGANISM: Unknown  
 FEATURE:  
 OTHER INFORMATION: Obtained from an environmental sample  
 US-10-146-772-57

Query Match 49.1%; Score 497.8; DB 6; Length 1014;  
 Best Local Similarity 68.2%; Pred. No. 5.9e-161;  
 Matches 691; Conservative 0; Mismatches 322; Indels 0; Gaps 0;

QY 2 TGAAGAAGCTATCAAGTCCCTGCGTCAAGCCGCCGATCTACATGATTTGGAGG 61  
 DB 2 TGAAGAAGCTATCAAGTCCCTGCGTCAAGCCGCCGATCTACATGATTTGGAGG 61  
 QY 62 CGAGCGTGAAGAAACATGATGATGATGATGATGATGATGATGATGATGATGATGAT 121  
 DB 62 CGAGCGTGAAGAAACATGATGATGATGATGATGATGATGATGATGATGATGATGAT 121  
 QY 62 CCAAGTGAAGAAACATGATGATGATGATGATGATGATGATGATGATGATGATGAT 121  
 DB 62 CCAAGTGAAGAAACATGATGATGATGATGATGATGATGATGATGATGATGATGAT 121  
 QY 122 TCGCCTTTCCGAAACCTTGATTCAGGCTACCCATGATGATGATGATGATGATGAT 181  
 DB 122 TCGCCTTTCCGAAACCTTGATTCAGGCTACCCATGATGATGATGATGATGATGAT 181  
 QY 182 CATGGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 241  
 DB 182 CATGGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 241  
 QY 182 CCGGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 241  
 DB 182 CCGGCGGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 241  
 QY 242 AAGCTAAGCGCATTTCAAGTGAAGCAAGCGGTTGGAATCATGATGATGATGATGAT 301  
 DB 242 AAGCTAAGCGCATTTCAAGTGAAGCAAGCGGTTGGAATCATGATGATGATGATGAT 301  
 QY 242 AGGCGAAGCGCATTCAGTGAAGCGCGCCAGCGCGGATATATACCTCCGCTAGGGTACA 301  
 DB 242 AGGCGAAGCGCATTCAGTGAAGCGCGCGCCAGCGCGGATATATACCTCCGCTAGGGTACA 301  
 QY 302 GTGAACGGGTGCGTGAAGCAACCTTTATCATGATGATGATGATGATGATGATGATGAT 361  
 DB 302 GTGAACGGGTGCGTGAAGCAACCTTTATCATGATGATGATGATGATGATGATGATGAT 361  
 QY 302 GCGAAGCGGTGCGTGAAGCAACCTTTATCATGATGATGATGATGATGATGATGATGAT 361  
 DB 302 GCGAAGCGGTGCGTGAAGCAACCTTTATCATGATGATGATGATGATGATGATGATGAT 361  
 QY 362 CCATGCGGCGCGCGGCAAGATGATGATGATGATGATGATGATGATGATGATGATGAT 421  
 DB 362 CCATGCGGCGCGCGGCAAGATGATGATGATGATGATGATGATGATGATGATGATGAT 421  
 QY 362 CAGCTGCGTGAAGCAAGCGGTTGGAATCATGATGATGATGATGATGATGATGATGAT 421  
 DB 362 CAGCTGCGTGAAGCAAGCGGTTGGAATCATGATGATGATGATGATGATGATGATGAT 421  
 QY 422 GCGATGCTTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 481  
 DB 422 GCGATGCTTATGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 481  
 QY 422 GCGACGATCATCCCTTTCCATTTGCAACACCGTTGCGGCTGCGGCACTTCTGCT 481  
 DB 422 GCGACGATCATCCCTTTCCATTTGCAACACCGTTGCGGCTGCGGCACTTCTGCT 481  
 QY 482 GTTGGAGAGACCTTCAACCGCTTAAACAAATACGCTTTGATGACAAATGAGAGATTC 541  
 DB 482 GTTGGAGAGACCTTCAACCGCTTAAACAAATACGCTTTGATGACAAATGAGAGATTC 541  
 QY 482 GTTGGAGAGACCTTCAACCGCTTAAACAAATACGCTTTGATGACAAATGAGAGATTC 541  
 DB 482 GTTGGAGAGACCTTCAACCGCTTAAACAAATACGCTTTGATGACAAATGAGAGATTC 541  
 QY 542 ATTGTCGCGTTGCGGAGCTTTAGCTTTATCTTATGCGGCAAGCCCTGCGGCTG 601  
 DB 542 ATTGTCGCGTTGCGGAGCTTTAGCTTTATCTTATGCGGCAAGCCCTGCGGCTG 601  
 QY 542 ACTTCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 601  
 DB 542 ACTTCGCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 601  
 QY 602 ATGTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 661  
 DB 602 ATGTCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 661  
 QY 602 AATTAATGTCGAGCTTCTGCGATCTACCGCTGAGAGGAGCTGTTTGTCTGCTGCT 661  
 DB 602 AATTAATGTCGAGCTTCTGCGATCTACCGCTGAGAGGAGCTGTTTGTCTGCTGCT 661  
 QY 662 CGTGTGCGCTGCTTCAACATTCATGATGATGATGATGATGATGATGATGATGATGATGAT 721  
 DB 662 CGTGTGCGCTGCTTCAACATTCATGATGATGATGATGATGATGATGATGATGATGATGAT 721  
 QY 662 CCGTGTGCGCTGCTTCAACATTCATGATGATGATGATGATGATGATGATGATGATGATGAT 721  
 DB 662 CCGTGTGCGCTGCTTCAACATTCATGATGATGATGATGATGATGATGATGATGATGATGAT 721  
 QY 722 CGTGTGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 781  
 DB 722 CGTGTGCTTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 781  
 QY 722 GCGTCTTCAAGCGCGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 781  
 DB 722 GCGTCTTCAAGCGCGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 781  
 QY 782 TCGGCGCTGCTTCAAGCGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 841  
 DB 782 TCGGCGCTGCTTCAAGCGCGGCGGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 841  
 QY 782 CGGCGCTTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 841  
 DB 782 CGGCGCTTGTGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 841

QY 842 GCATCCTTGTAAATGCGCGGAGACCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 901  
 DB 842 GCATCCTTGTAAATGCGCGGAGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 901  
 QY 902 GCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 961  
 DB 902 GCTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 961  
 QY 962 CTTACGCTTTGGGTTAAGCGCTGAGAGCGGCTGCGGCACTCGAAGAAATTTGA 1014  
 DB 962 CTTACGCTTTGGGTTAAGCGCTGAGAGCGGCTGCGGCACTCGAAGAAATTTGA 1014  
 QY 962 CATCGGCGCGGCGGCTTTGAAAGTTGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1014  
 DB 962 CATCGGCGCGGCGGCTTTGAAAGTTGAGGCGGCGGCGGCGGCGGCGGCGGCGGCGG 1014

RESULT 13  
 US-10-241-742-57  
 Sequence 57, Application US/10241742  
 Publication No. US20040002147A1  
 GENERAL INFORMATION:  
 APPLICANT: Short, Jay  
 APPLICANT: Weiner, David  
 APPLICANT: Chaplin, Jennifer  
 APPLICANT: Chli, Ellen  
 APPLICANT: Milam, Aileen  
 APPLICANT: Desantis, Grace  
 APPLICANT: Madden, Mark  
 APPLICANT: Burk, Mark  
 TITLE OR INVENTION: Nitrilases  
 FILE REFERENCE: Docket No. US20040002147A1 DIV-013US  
 CURRENT APPLICATION NUMBER: US/10/241,742  
 PRIOR FILING DATE: 2002-09-09  
 PRIOR APPLICATION NUMBER: US 60/146,772  
 PRIOR FILING DATE: 2002-05-15  
 PRIOR APPLICATION NUMBER: US 60/309,006  
 PRIOR FILING DATE: 2001-07-30  
 PRIOR APPLICATION NUMBER: US 60/351,336  
 PRIOR FILING DATE: 2002-01-22  
 PRIOR APPLICATION NUMBER: US 60/300,189  
 PRIOR FILING DATE: 2001-06-21  
 PRIOR APPLICATION NUMBER: US 09/751,299  
 PRIOR FILING DATE: 2000-12-28  
 PRIOR APPLICATION NUMBER: US 60/254,414  
 PRIOR FILING DATE: 2000-12-07  
 PRIOR APPLICATION NUMBER: US 60/173,609  
 PRIOR FILING DATE: 1999-12-29  
 NUMBER OF SEQ ID NOS: 386  
 SOFTWARE: FastSeq for Windows Version 4.0  
 SEQ ID NO 57  
 LENGTH: 1014  
 TYPE: DNA  
 ORGANISM: Unknown  
 FEATURE:  
 OTHER INFORMATION: Obtained from an environmental sample  
 US-10-241-742-57

Query Match 49.1%; Score 497.8; DB 6; Length 1014;  
 Best Local Similarity 68.2%; Pred. No. 5.9e-161;  
 Matches 691; Conservative 0; Mismatches 322; Indels 0; Gaps 0;

QY 2 TGAAGAAGCTATCAAGTCCCTGCGTCAAGCCGCCGATCTACATGATTTGGAGG 61  
 DB 2 TGAAGAAGCTATCAAGTCCCTGCGTCAAGCCGCCGATCTACATGATTTGGAGG 61  
 QY 62 CGAGCGTGAAGAAACATGATGATGATGATGATGATGATGATGATGATGATGATGAT 121  
 DB 62 CGAGCGTGAAGAAACATGATGATGATGATGATGATGATGATGATGATGATGATGAT 121  
 QY 62 CCAAGTGAAGAAACATGATGATGATGATGATGATGATGATGATGATGATGATGAT 121  
 DB 62 CCAAGTGAAGAAACATGATGATGATGATGATGATGATGATGATGATGATGATGAT 121  
 QY 122 TCGCCTTTCCGAAACCTTGATTCAGGCTACCCATGATGATGATGATGATGATGATGAT 181  
 DB 122 TCGCCTTTCCGAAACCTTGATTCAGGCTACCCATGATGATGATGATGATGATGATGAT 181  
 QY 182 CATGGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 241  
 DB 182 CATGGCAATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 241

```
Db 182 CCTGGGGATGCAATTCGTGCGCCGATACCAAGAACTCACTGGTCTCGACAGCCCTC 241
Qy 242 AAGCTAAGCGCATTTTCAGATGAGCGCGGTTGGAAATCATGTGATCCCTGGGGATGA 301
Db 242 AGGCCAAGGCGATCAGTGAAGCGCGCCAGCGCGGTATATATGCTGGGCTAGGGATCA 301
Qy 302 GTGAACGGGTCCGGTGGCACTTTTATCATGATCAGTGGTTCATAGGCGATTAATGTGACA 361
Db 302 GGGAAAGCGGTGAGCGGAACCTCTACATGGGGCAATGGGTCTATTGACATTAAGGCGAAA 361
Qy 362 CCATTGGGGCCCGCGGAAAGTTGAAACCTACTTTTGTGGAAGCTACTTTGTGGCGAAG 421
Db 362 CAGCTGGGCGTGGCGGAAAGCTGAAACCAACCATGATGAGGAAACCTCTTCGGTGAAG 421
Qy 422 GGGATGTTTCATGCGGTAGCGGTTTCGAGACGTCGTGGTGAAGGCTGGGTGGCTTATGT 481
Db 422 GCGACGAGATCATCCCTTTTCACCTTTCGACACCGTTGGGGGTGCTGGGCGGACCTGTG 481
Qy 482 GTTGGGAGCACCTTTCACCGCTAACAAATACGTTTGTATGACAAATGAAAGATTC 541
Db 482 GTTGGGAACTTACCACTCTTTGAAATATGCGCTCTACGACAGAAAGTCTTGGACCG 541
Qy 542 ATTTGGCGGCTTGGCGGAGCTTTTAACTTATCTTAATGCGGCGAAAGCCCTGGGGCCTG 601
Db 542 ACTTGGCGGCTGGCGCTAGCTTACGATCTACGTCAGCGACGACGAAAGTCTTGGACCG 601
Qy 602 ATGTCAATGTAAGCGGCTCTGCAATCTATGCGGTTGAAGGGCAATGCTTCTACTAGGT 661
Db 602 AAGTAAATGTGCGACCTTCTCGAATCTACGCGTGAAGGGCAATGTTTGTCTCGCTT 661
Qy 662 CGTGGCGGCTGTTTCACAATCATGATGATGATGATGATGATGATGATGATGATGATGATG 721
Db 662 CCTGGCGGCTGCTCTGCGACGATGATGATGATGATGATGATGATGATGATGATGATG 721
Qy 722 GCTTCTTCAAGGCGCGCGGCGGCTTCTCCGCAATTATGGTCCGATGGACGACGCTAG 781
Db 722 GCTTCTTCAAGGCGCGCGGCGGCTTCTCCGCAATTATGGTCCGATGGACGACGCTAG 781
Qy 782 TCGCGGCTCTTGGCGGAAATGAAGGGTATCTCTAGGCAAACTTGTATCTTGAAGTAC 841
Db 782 CGCGGCGCTTGGCGGAAATGAAGGGTATCTCTAGGCAAACTTGTATCTTGAAGTAC 841
Qy 842 GCATCCTTGTAAATGAGGAGACCGCTGCTGATCATTAATCCGCTCCGACATTAATC 901
Db 842 GAATTAATGCAAAAGCGGAGCTGATCAAGCGGCGCATCTCCGACGACGATCACTC 901
Qy 902 GCTTGTAAATGATGCAAGCGCTTAATTAACCGGTAGTTGAAATTTGAAGGTATCTTCTG 961
Db 902 GCTTGTAAATGATGCAAGCGCTTAATTAACCGGTAGTTGAAATTTGAAGGTATCTTCTG 961
Qy 962 CTTTACGCTTTGGGTAAAGCGTCTGAGAGGGGTGCGCAACTGGAAGAAATTTGA 1014
Db 962 CATCGGCCCAAGGCTTTGAAAGTTGAGGGGCGGCCCGGGGTACGAGGCGCATTTGA 1014
```

RESULT 14  
US-10-440-523-57

```
; Sequence 57, Application US/10440523
; Publication No. US20040014195A1
; GENERAL INFORMATION:
; APPLICANT: Short, Jay
; APPLICANT: Weiner, David
; APPLICANT: Chaplin, Jennifer
; APPLICANT: Chi, Ellen
; APPLICANT: Milan, Aileen
; APPLICANT: Desantis, Grace
; APPLICANT: Madden, Mark
; APPLICANT: Burk, Mark
; TITLE OF INVENTION: Nitric Oxide
; FILE REFERENCE: Docket No. US20040014195A1 DIV-013US
; CURRENT APPLICATION NUMBER: US/10/440,523
; CURRENT FILING DATE: 2003-05-15
; PRIOR APPLICATION NUMBER: US/10/146,772
```

```
; PRIOR FILING DATE: 2002-05-15
; PRIOR APPLICATION NUMBER: US 60/309,006
; PRIOR FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US 60/351,336
; PRIOR FILING DATE: 2002-01-22
; PRIOR APPLICATION NUMBER: US 60/300,189
; PRIOR FILING DATE: 2001-06-21
; PRIOR APPLICATION NUMBER: US 09/751,299
; PRIOR FILING DATE: 2000-12-28
; PRIOR APPLICATION NUMBER: US 60/254,414
; PRIOR FILING DATE: 2000-12-07
; PRIOR APPLICATION NUMBER: US 60/173,609
; PRIOR FILING DATE: 1999-12-29
; NUMBER OF SEQ ID NOS: 386
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 57
; LENGTH: 1014
; TYPE: DNA
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Obtained from an environmental sample
; US-10-440-523-57

Query Match 49.1%; Score 497.8; DB 7; Length 1014;
Best Local Similarity 68.2%; Pred. No. 5.9e-161;
Matches 691; Conservative 0; Mismatches 322; Indels 0; Gaps 0;

Qy 2 TGAAGAAAGTATCAAGTCCCTGCGTGAAGCCCGCCGATCTACATGATTTGAGG 61
Db 2 TGAAGAAAGTATCAAGTCCCTGCGTGAAGCCCGCCGATCTACATGATTTGAGG 61
Qy 62 CGACGCTGGAAGAAACCTTGAATGAGTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 121
Db 62 CGACGCTGGAAGAAACCTTGAATGAGTGAAGGAGGAGGAGGAGGAGGAGGAGGAGG 121
Qy 122 TCGCCTTCCGGAACCTTGAATGAGTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 181
Db 122 TCGCCTTCCGGAACCTTGAATGAGTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 181
Qy 122 TCGCCTTCCGGAACCTTGAATGAGTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 181
Db 122 TCGCCTTCCGGAACCTTGAATGAGTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 181
Qy 182 CATGGGCAATGCAATTTGTACGCGCAATACATGAGAACTCATTTGATGAGTGGATG 241
Db 182 CATGGGCAATGCAATTTGTACGCGCAATACATGAGAACTCATTTGATGAGTGGATG 241
Qy 182 CCGTGGGAGTCAATTTGCGCGGATGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 241
Db 182 CCGTGGGAGTCAATTTGCGCGGATGATGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 241
Qy 242 AAGCTAAGCGATTTGCAAGTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 301
Db 242 AAGCTAAGCGATTTGCAAGTGAAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 301
Qy 302 GTGAACGGGTCCGGTGGCACTTTTATCATGATCAGTGGTTCATAGGCGATTAATGTGACA 361
Db 302 GTGAACGGGTCCGGTGGCACTTTTATCATGATCAGTGGTTCATAGGCGATTAATGTGACA 361
Qy 302 GCGAAGCGGTGAGCGGAACCTCTACATGGGGCAATGGGTCTATTGACATTAAGGCGAAA 361
Db 302 GCGAAGCGGTGAGCGGAACCTCTACATGGGGCAATGGGTCTATTGACATTAAGGCGAAA 361
Qy 362 CCATTGGGGCCCGCGGAAAGTTGAAACCTACTTTTGTGGAAGCTACTTTGTGGCGAAG 421
Db 362 CCATTGGGGCCCGCGGAAAGTTGAAACCTACTTTTGTGGAAGCTACTTTGTGGCGAAG 421
Qy 422 GGGATGTTTCATGCGGTAGCGGTTTCGAGACGTCGTGGTGAAGGCTGGGTGGCTTATGT 481
Db 422 GGGATGTTTCATGCGGTAGCGGTTTCGAGACGTCGTGGTGAAGGCTGGGTGGCTTATGT 481
Qy 422 GCGACGAGATCATCCCTTTTCACCTTTCGACACCGTTGGGGGTGCTGGGCGGACCTGTG 481
Db 422 GCGACGAGATCATCCCTTTTCACCTTTCGACACCGTTGGGGGTGCTGGGCGGACCTGTG 481
Qy 482 GTTGGGAGCACCTTTCACCGCTAACAAATACGTTTGTATGACAAATGAAAGATTC 541
Db 482 GTTGGGAGCACCTTTCACCGCTAACAAATACGTTTGTATGACAAATGAAAGATTC 541
Qy 482 GTTGGGAACTTACCACTCTTTGAAATATGCGCTCTACGACAGAAAGTCTTGGACCG 541
Db 482 GTTGGGAACTTACCACTCTTTGAAATATGCGCTCTACGACAGAAAGTCTTGGACCG 541
Qy 542 ATTTGGCGGCTTGGCGGAGCTTTTAACTTATCTTAATGCGGCGAAAGCCCTGGGGCCTG 601
Db 542 ATTTGGCGGCTTGGCGGAGCTTTTAACTTATCTTAATGCGGCGAAAGCCCTGGGGCCTG 601
Qy 542 ACTTGGCGGCTGGCGCTAGCTTACGATCTACGTCAGCGACGACGAAAGTCTTGGACCG 601
Db 542 ACTTGGCGGCTGGCGCTAGCTTACGATCTACGTCAGCGACGACGAAAGTCTTGGACCG 601
Qy 602 ATGTCAATGTAAGCGGCTCTGCAATCTATGCGGTTGAAGGGCAATGCTTCTACTAGGT 661
Db 602 ATGTCAATGTAAGCGGCTCTGCAATCTATGCGGTTGAAGGGCAATGCTTCTACTAGGT 661
Qy 602 AAGTAAATGTGCGACCTTCTCGAATCTACGCGTGAAGGGCAATGTTTGTCTCGCTT 661
Db 602 AAGTAAATGTGCGACCTTCTCGAATCTACGCGTGAAGGGCAATGTTTGTCTCGCTT 661
Qy 662 CGTGGCGGCTGTTTCACAATCATGATGATGATGATGATGATGATGATGATGATGATGATG 721
Db 662 CGTGGCGGCTGTTTCACAATCATGATGATGATGATGATGATGATGATGATGATGATGATG 721
```



GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM nucleic - nucleic search, using sw model

Run on: April 26, 2006, 23:27:34 ; Search time 609.88 Seconds  
(without alignments)  
6752.353 Million cell updates/sec

Title: US-09-751-299-3

Perfect score: 1014

Sequence: 1 atgaagaagcgtcctcaaggt.....cgcaactcgagaatttga 1014

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 9295968 seqs, 2030634719 residues

Total number of hits satisfying chosen parameters: 18591936

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

Published Applications NA.New:\*  
1: /SID55/prodata/2/pubpna/US08\_NEW\_PUB.seq:\*  
2: /SID55/prodata/2/pubpna/US06\_NEW\_PUB.seq:\*  
3: /SID55/prodata/2/pubpna/US07\_NEW\_PUB.seq:\*  
4: /SID55/prodata/2/pubpna/PCT\_NEW\_PUB.seq:\*  
5: /SID55/prodata/2/pubpna/US09\_NEW\_PUB.seq1:\*  
6: /SID55/prodata/2/pubpna/US10\_NEW\_PUB.seq1:\*  
7: /SID55/prodata/2/pubpna/US10\_NEW\_PUB.seq2:\*  
8: /SID55/prodata/2/pubpna/US10\_NEW\_PUB.seq2:\*  
9: /SID55/prodata/2/pubpna/US10\_NEW\_PUB.seq3:\*  
10: /SID55/prodata/2/pubpna/US10\_NEW\_PUB.seq3:\*  
11: /SID55/prodata/2/pubpna/US11\_NEW\_PUB.seq2:\*  
12: /SID55/prodata/2/pubpna/US11\_NEW\_PUB.seq2:\*  
13: /SID55/prodata/2/pubpna/US11\_NEW\_PUB.seq3:\*  
14: /SID55/prodata/2/pubpna/US11\_NEW\_PUB.seq4:\*  
15: /SID55/prodata/2/pubpna/US60\_NEW\_PUB.seq:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	248.6	24.5	1071	9	US-10-537-075-6 Sequence 6, Appl1
2	219.6	21.7	1110	9	US-10-919-182-15 Sequence 15, Appl1
3	219.6	21.7	1110	9	US-10-919-182-17 Sequence 17, Appl1
4	218	21.5	1110	9	US-10-919-182-3 Sequence 3, Appl1
5	218	21.5	1110	9	US-10-919-182-5 Sequence 5, Appl1
6	218	21.5	1110	9	US-10-919-182-11 Sequence 11, Appl1
7	216.4	21.3	1110	9	US-10-919-182-7 Sequence 7, Appl1
8	216.4	21.3	1110	9	US-10-919-182-13 Sequence 13, Appl1
9	83	8.2	1415	11	US-11-096-568A-6953 Sequence 6953, Ap
10	81.6	8.0	1706	11	US-11-096-568A-20685 Sequence 20685, A
11	70.4	6.9	597	9	US-10-932-182A-166788 Sequence 166788,
12	70.4	6.9	597	9	US-10-932-182A-166788 Sequence 166788,
13	34.6	3.4	548	6	US-09-925-065A-628387 Sequence 628387,
14	34	3.4	548	6	US-09-925-065A-628387 Sequence 628387,
15	34	3.4	563	10	US-10-301-480-489364 Sequence 489364,
16	33.6	3.3	579	9	US-10-301-480-56141 Sequence 56141, A
17	33.6	3.3	579	9	US-10-301-480-56141 Sequence 56141, A
18	33.6	3.3	579	10	US-10-301-480-669550 Sequence 669550,

19	33.2	3.3	3815	14	US-11-099-855-1 Sequence 1, Appl1
20	33	3.3	42379	9	US-10-330-773-637 Sequence 637, App
21	32.8	3.2	609	9	US-10-330-773-637 Sequence 75064, A
22	32.8	3.2	609	10	US-10-301-480-688473 Sequence 688473,
23	32.8	3.2	2842	10	US-10-644-807-52 Sequence 52, Appl1
24	32.6	3.2	336	8	US-10-623-155-286 Sequence 286, App
25	32.4	3.2	4620	14	US-11-136-557-1028 Sequence 3028, Ap
26	32	3.2	1218	8	US-10-750-185-64760 Sequence 64760, A
27	32	3.2	1218	8	US-10-750-623-64760 Sequence 64760, A
28	31.6	3.1	574	6	US-09-925-065A-368038 Sequence 368038,
29	31.6	3.1	576	10	US-10-301-480-438304 Sequence 438304,
30	31.6	3.1	576	10	US-10-301-480-1051713 Sequence 1051713,
31	31.6	3.1	984	10	US-10-301-480-531203 Sequence 531203,
32	31.6	3.1	984	10	US-10-301-480-1144612 Sequence 1144612,
33	31.4	3.1	621	10	US-10-301-480-475097 Sequence 475097,
34	31.4	3.1	621	10	US-10-301-480-1088506 Sequence 1088506,
35	31.4	3.1	622	6	US-09-925-065A-408575 Sequence 408575,
36	31.4	3.1	1471	11	US-11-183-327-1 Sequence 1, Appl1
37	31.4	3.1	1471	11	US-11-183-327-24 Sequence 24, Appl1
38	31.4	3.1	1702	11	US-11-072-512-919 Sequence 919, App
39	31.4	3.1	1905	11	US-11-072-512-1070 Sequence 1070, Ap
40	31.4	3.1	3289	8	US-10-750-185-31224 Sequence 31224, A
41	31.4	3.1	3289	8	US-10-750-623-31224 Sequence 31224, A
42	31.4	3.1	3612	6	US-09-925-065A-702369 Sequence 702369,
43	31.4	3.1	4053	11	US-11-072-512-539 Sequence 539, App
44	31.2	3.1	513	6	US-09-925-065A-187799 Sequence 187799,
45	31.2	3.1	513	6	US-09-925-065A-187800 Sequence 187800,

#### ALIGNMENTS

RESULT 1	US-10-537-075-6	Sequence 6, Application US/10537075
1	US-10-537-075-6	Sequence 6, Application US/10537075
2	US-10-537-075-6	Sequence 6, Application US/10537075
3	US-10-537-075-6	Sequence 6, Application US/10537075
4	US-10-537-075-6	Sequence 6, Application US/10537075
5	US-10-537-075-6	Sequence 6, Application US/10537075
6	US-10-537-075-6	Sequence 6, Application US/10537075
7	US-10-537-075-6	Sequence 6, Application US/10537075
8	US-10-537-075-6	Sequence 6, Application US/10537075
9	US-10-537-075-6	Sequence 6, Application US/10537075
10	US-10-537-075-6	Sequence 6, Application US/10537075
11	US-10-537-075-6	Sequence 6, Application US/10537075
12	US-10-537-075-6	Sequence 6, Application US/10537075
13	US-10-537-075-6	Sequence 6, Application US/10537075
14	US-10-537-075-6	Sequence 6, Application US/10537075
15	US-10-537-075-6	Sequence 6, Application US/10537075
16	US-10-537-075-6	Sequence 6, Application US/10537075
17	US-10-537-075-6	Sequence 6, Application US/10537075
18	US-10-537-075-6	Sequence 6, Application US/10537075
19	US-10-537-075-6	Sequence 6, Application US/10537075
20	US-10-537-075-6	Sequence 6, Application US/10537075
21	US-10-537-075-6	Sequence 6, Application US/10537075
22	US-10-537-075-6	Sequence 6, Application US/10537075
23	US-10-537-075-6	Sequence 6, Application US/10537075
24	US-10-537-075-6	Sequence 6, Application US/10537075
25	US-10-537-075-6	Sequence 6, Application US/10537075
26	US-10-537-075-6	Sequence 6, Application US/10537075
27	US-10-537-075-6	Sequence 6, Application US/10537075
28	US-10-537-075-6	Sequence 6, Application US/10537075
29	US-10-537-075-6	Sequence 6, Application US/10537075
30	US-10-537-075-6	Sequence 6, Application US/10537075
31	US-10-537-075-6	Sequence 6, Application US/10537075
32	US-10-537-075-6	Sequence 6, Application US/10537075
33	US-10-537-075-6	Sequence 6, Application US/10537075
34	US-10-537-075-6	Sequence 6, Application US/10537075
35	US-10-537-075-6	Sequence 6, Application US/10537075
36	US-10-537-075-6	Sequence 6, Application US/10537075
37	US-10-537-075-6	Sequence 6, Application US/10537075
38	US-10-537-075-6	Sequence 6, Application US/10537075
39	US-10-537-075-6	Sequence 6, Application US/10537075
40	US-10-537-075-6	Sequence 6, Application US/10537075
41	US-10-537-075-6	Sequence 6, Application US/10537075
42	US-10-537-075-6	Sequence 6, Application US/10537075
43	US-10-537-075-6	Sequence 6, Application US/10537075
44	US-10-537-075-6	Sequence 6, Application US/10537075
45	US-10-537-075-6	Sequence 6, Application US/10537075
46	US-10-537-075-6	Sequence 6, Application US/10537075
47	US-10-537-075-6	Sequence 6, Application US/10537075
48	US-10-537-075-6	Sequence 6, Application US/10537075
49	US-10-537-075-6	Sequence 6, Application US/10537075
50	US-10-537-075-6	Sequence 6, Application US/10537075
51	US-10-537-075-6	Sequence 6, Application US/10537075
52	US-10-537-075-6	Sequence 6, Application US/10537075
53	US-10-537-075-6	Sequence 6, Application US/10537075
54	US-10-537-075-6	Sequence 6, Application US/10537075
55	US-10-537-075-6	Sequence 6, Application US/10537075
56	US-10-537-075-6	Sequence 6, Application US/10537075
57	US-10-537-075-6	Sequence 6, Application US/10537075
58	US-10-537-075-6	Sequence 6, Application US/10537075
59	US-10-537-075-6	Sequence 6, Application US/10537075
60	US-10-537-075-6	Sequence 6, Application US/10537075
61	US-10-537-075-6	Sequence 6, Application US/10537075
62	US-10-537-075-6	Sequence 6, Application US/10537075
63	US-10-537-075-6	Sequence 6, Application US/10537075
64	US-10-537-075-6	Sequence 6, Application US/10537075
65	US-10-537-075-6	Sequence 6, Application US/10537075
66	US-10-537-075-6	Sequence 6, Application US/10537075
67	US-10-537-075-6	Sequence 6, Application US/10537075
68	US-10-537-075-6	Sequence 6, Application US/10537075
69	US-10-537-075-6	Sequence 6, Application US/10537075
70	US-10-537-075-6	Sequence 6, Application US/10537075
71	US-10-537-075-6	Sequence 6, Application US/10537075
72	US-10-537-075-6	Sequence 6, Application US/10537075
73	US-10-537-075-6	Sequence 6, Application US/10537075
74	US-10-537-075-6	Sequence 6, Application US/10537075
75	US-10-537-075-6	Sequence 6, Application US/10537075
76	US-10-537-075-6	Sequence 6, Application US/10537075
77	US-10-537-075-6	Sequence 6, Application US/10537075
78	US-10-537-075-6	Sequence 6, Application US/10537075
79	US-10-537-075-6	Sequence 6, Application US/10537075
80	US-10-537-075-6	Sequence 6, Application US/10537075
81	US-10-537-075-6	Sequence 6, Application US/10537075
82	US-10-537-075-6	Sequence 6, Application US/10537075
83	US-10-537-075-6	Sequence 6, Application US/10537075
84	US-10-537-075-6	Sequence 6, Application US/10537075
85	US-10-537-075-6	Sequence 6, Application US/10537075
86	US-10-537-075-6	Sequence 6, Application US/10537075
87	US-10-537-075-6	Sequence 6, Application US/10537075
88	US-10-537-075-6	Sequence 6, Application US/10537075
89	US-10-537-075-6	Sequence 6, Application US/10537075
90	US-10-537-075-6	Sequence 6, Application US/10537075
91	US-10-537-075-6	Sequence 6, Application US/10537075
92	US-10-537-075-6	Sequence 6, Application US/10537075
93	US-10-537-075-6	Sequence 6, Application US/10537075
94	US-10-537-075-6	Sequence 6, Application US/10537075
95	US-10-537-075-6	Sequence 6, Application US/10537075
96	US-10-537-075-6	Sequence 6, Application US/10537075
97	US-10-537-075-6	Sequence 6, Application US/10537075
98	US-10-537-075-6	Sequence 6, Application US/10537075
99	US-10-537-075-6	Sequence 6, Application US/10537075
100	US-10-537-075-6	Sequence 6, Application US/10537075

```
Db 130 GTGTTGGTGAACCTGGCTGCGCGATATCCCTTCCACGCTCTGGCTGGCGCACCGGCC 189
Qy 184 TGGGCATGCAATTTGTAGCCCAATACCATGAGAGATGAGTTGAGTTGAGTGGCCCTCA 243
Db 190 TGGTGGCTGAAATACAGTCCCGCTACTATGCAACTGCTCTCGCTGACAGTGCAGAG 249
Qy 244 GCTAAGCCGATTTTCAGATGACAGCCAGCGGTTGGAAATCATAGTCACTTGGGGATGAGT 303
Db 250 TTTCAACCCATTTGGCCAGCGCACGAGACCTTGGGTAATTTTCATGCACTGGGTTATAGC 309
Qy 304 GAACGGGTGCGTGGACCCCTTACATCATGTCAGTGGTTCATATGAGGATATATGATACCC 363
Db 310 GAGCCACAGCGCGAGGCTTTTACCTGGGCCAATCCCTGATGACAGACAGGCGAGATG 369
Qy 364 ATTGGGCGCCGCGAAAGTTGAACCTTATCTTTGTTGAACGTAATTTGTTGGCGAGAGG 423
Db 370 CTGTGGTGGCGTGCACAACTCAACCCACGCAATGAGAGCGCACCGTATTGGTGAAGGT 429
Qy 424 GATGTTATCGCTAGCGGTTTTCAGAGACGTCGTTGGAAGGCTGGGCTTATAGCTGT 483
Db 430 TATGCCGCTGATCTGATTTGTGTCGACACAGAACTGGAGCGCTGGCTCTATGCTGC 489
Qy 484 TGGGAGCACTTTCACCCGCTAACAAATACGCTTTGTATGACAAATGAAAGATTTGAT 543
Db 490 TGGGAGCAATTTGTGCGCCCTTGAGCAAGTACGCGCTGTACTCCAGACATGAAGCCATTAC 549
Qy 544 TGTGCGGCTTGGCGGAGCTTTAGCCTTATCTTATCTTAATGCGGCGAAAGCCCTGGGGCTGAT 603
Db 550 ATTGCTGCTGGCGCTGCTGTTTCGCTATACAGCGAAACGCGCCACGCTCAGTCCAG 609
Qy 604 GTCAATGATAGCGGCTCTGCAATCTATGCGGTTGAAAGGCAATGCTTGTATACGTCG 663
Db 610 GTGAACATGCGCTGCTCGCAATCTATTCGTTGAAGCGCAGTGTCTTACCATGCGGCC 669
Qy 664 TGTGCGCTGTTTCAACATTCATGATCGATATGCTTTGTATGACATGACGAAAGCATGCG 723
Db 670 AGCACTGTGCTACCCCAAGAGACGCTAGACATGCTGAAGTGGTGAACACAAACCCGCC 729
Qy 724 TTGCTTGGGCTGTGTGTGAGACATCAGATCATAGGCGCTGATGAGTGTACTTGTGTC 783
Db 730 TTGCTGAAAGTGGGCGGCGAGTTCATGATTTTGGCGCGGACGAGACATGCTGCT 789
Qy 784 GCGCCTCTTCCGAAATGAAAGAGGTATCTCTACGCAACCTTGATCTGAGTACGC 843
Db 790 CCTACCTGCTCAGATGCGGAGGCTTGATCATTTGCCGATCTGAATGAGAGATTT 849
Qy 844 ATCTTGTAAATGCGCGGAGACCTGCTGCTCATTTTCCGTCGCGACATTTACTGCG 903
Db 850 GCTTTCGCAAGCATCATATGACCCCGAGGCACTATTCGAAACCGGAGCCACCGCT 909
Qy 904 TTGCTAATAGA 914
Db 910 CTGGTGTGGA 920

RESULT 2
US-10-919-182-15
; Sequence 15, Application US/10919182
; Publication No. US2006003532A1
; GENERAL INFORMATION:
; APPLICANT: E. I. duPont de Nemours and Company, Inc.
; APPLICANT: Di Cosimo, Robert
; APPLICANT: Payne, Mark
; APPLICANT: O'Keefe, Daniel
; TITLE OF INVENTION: PRODUCTION OF 3-HYDROXYCARBOXYLIC ACID USING NITRILASE MUTANTS
; FILE REFERENCE: CL2584 US NA
; CURRENT APPLICATION NUMBER: US/10/919,182
; CURRENT FILING DATE: 2004-08-16
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 15
; LENGTH: 1110
```

```
; TYPE: DNA
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Mutant nitrilase with Phe168 to Val change
; NAME/KEY: CDS
; LOCATION: (1)..(1110)
US-10-919-182-15

Query Match 21.7%; Score 219.6; DB 9; Length 1110;
Best Local Similarity 52.1%; Pred. No. 1.1e-60;
Matches 489; Conservative 0; Mismatches 449; Indels 0; Gaps 0;

Qy 22 GCGTGGCGAAGCGCCCGCATCTACATGGAATTTGGAGGCGAGCGTGAACAAACCTT 81
Db 31 GCAACGTTACGAGAGAGCGGTATGGCTGACGCAAGCGCAACGATCGACAGTTCATC 90
Qy 82 GAGTTGATGGAAGAAGACGACGTAATATGCTGCTGATCGCCTTTCGGAACCTTGG 141
Db 91 GGCATCATCGAAGAAAGGTGCCCAAAAGGCGCGAGTCTGATCGCTTCCGGAATATTC 150
Qy 142 ATTCCAGGCTACCAATGTTCTTTGGCTTACCTACACAGCATGGCAATGCAATTTGTA 201
Db 151 ATTCCGGCTACCCCTATTGCGCGTGGCTGCGACAGTGAAGTACAGCTTAAGCTTACT 210
Qy 202 CGCAATACCATGAGAACTATTGAGTGGATGAGCCCTCAAGCTTAAGCGCATTTCAAGT 261
Db 211 TCAGGCTATCAGAAATTTGTTGAGCTAGGTACACACGTAAGCGTCTGCTCAGCTG 270
Qy 262 GCAGCCAAAGCGTTGGGAATCATGATCACTTGGGGATGATGAACGGGTGCTGGCAC 321
Db 271 GCCCGCGCCCAACAAATGCACTGTCATGCGCTATTTCGAGCGGGAAGCCGGATTCG 330
Qy 322 CTTATACATCAGTGTGTTGATAGGCGATATAGTGTGACATTTGGGCGCGGAAAG 381
Db 331 CGCTATCTGACCCAGGTGTTATGACGAGCGGTGCGAAGATCTGTTGCCAATCGCGCAAG 390
Qy 382 TTGAAACCTACTTTTGTGAACGTAATCTTGTTCGCGAAGGGAGTGTCAATCGCTACG 441
Db 391 CTGAAGCCCAACACAGTTGAGCGTACGATCTACGCGGCAAGCAACGAAATTCCTC 450
Qy 442 GTTTTCAGAGCTGTGTTGGAAGGCTGGGTGGCTTATGCTGTGGAGCACTTCAACG 501
Db 451 ACGACGATTCGCGTTCGACGCGTGGTGAATTAATCTCTGGAAACATGTTCAACG 510
Qy 502 CTAAACAAATACGCTTGTATGACAAATATGAAGATTTCTTGTGCGGCTTGGCCAGC 561
Db 511 CTCAGCAAGTTCAATGATGTACAGCTCGGTGAGCAGGTCCACGTTGCAATCGTGGCGG 570
Qy 562 TTTAGCCTTATTCCTAATGCGCGAAGACCTTGGGCGCTGATGTCAATGTAGCGGCTCT 621
Db 571 ATGTCCCTCTTACGCGCGATGTTTTCCAACTGTAGCATGAAAGCCAGCGAGTCAAC 630
Qy 622 CGAATCTATGCGGTTGAAGGCAATGCTTCTTACTAGCGTGTGCTGCTGTTTCAAA 681
Db 631 CGCTGTAACGAATGAAAGGCCAAACCTTTGTGCTTGTGCAAGCAGTATGAGCACT 690
Qy 682 TCATGATGATATGCTTTGTACAGATGACGAAAGCATGCGTGTCTTGTGCTGTGCTGT 741
Db 691 AGCGGATCGAAAGCTTGTGCTCAACGACGAAACGCGGACGCTGTTCGCCAGAGATGT 750
Qy 742 GGAACATGATATCATATGAGGCTGATGAGTGTGCTTGTGCGGCTTGGCCGAAAT 801
Db 751 GCGTGGGCGCATTTTACGCGCCCGAGATGAAGCGAGCTTGGAAAGCTTGGCGGAAT 810
Qy 802 GAAGAGGATTTCTCTACGCAACCTTGATCTGAGTACGATCTTGTCTAAATGCGG 861
Db 811 GGTAGGAGATCTTGTAGCAGAGATGATCTGAGCAGATTTGCTGGCAAGGCTGGA 870
Qy 862 GGAACCTGCTGTGATATTTCCGCTCCGACATTAATCTGCTGTCTAATAGATGCGAC 921
Db 871 GCCGATCGGATCGGACATTTGCGGCGCTACGCTGTGCTGATGATGACCGCGCG 930
```



Qy 922 CCTAATTACCGGTAGTTGAATTGAAGGTGATCTTCG 959  
Db 931 AATCATACGCCAGTTCATCGCATCGCATTAAGCGTCG 968

RESULT 3  
US-10-919-182-17

; Sequence 17, Application US/10919182  
; Publication No. US20060035352A1  
; GENERAL INFORMATION:  
; APPLICANT: E.I. duPont de Nemours and Company, Inc.  
; APPLICANT: Di Cosimo, Robert  
; APPLICANT: Payne, Mark  
; APPLICANT: O'Keefe, Daniel  
; TITLE OF INVENTION: PRODUCTION OF 3-HYDROXYCARBOXYLIC ACID USING NITRILASE MUTANTS  
; FILE REFERENCE: CL/584 US NA  
; CURRENT APPLICATION NUMBER: US/10/919,182  
; CURRENT FILING DATE: 2004-08-16  
; NUMBER OF SEQ ID NOS: 18  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO: 17  
; LENGTH: 1110  
; TYPE: DNA  
; ORGANISM: Artificial sequence  
; FEATURE:  
; OTHER INFORMATION: Mutant nitrilase with Phe168 to Leu change  
; NAME/KEY: CDS  
; LOCATION: (1)..(1110)  
US-10-919-182-17

Query Match 21.7%; Score 219.6; DB 9; Length 1110;

Best Local Similarity 52.1%; Pred. No. 1.1e-60; Indels 0; Gaps 0;

Matches 489; Conservative 0; Mismatches 449; Indels 0; Gaps 0;

Qy 22 GCCTCGTGCAGAGCGCCCGATCTACATGATTTGGAGCGCGGTGAGCAAAACATT 81  
Db 31 GCAACCGTTGACGAGAGCGCGGTATGCTCGACGACGACGACATGACAAAGTCGATC 90  
Qy 82 GAGTTGATGAAGAAGACGACGATTAATATGCTGCTGATGCGCTTTCCGAAACTTGG 141  
Db 91 GGCATCATCGAAGAAGCTGCCAAAGGCGCGAGTGTATGCTTTCCGAAAGTATTC 150  
Qy 142 ATTCAGGTACCCATGTTTCTTTGGCTTGACTACACGACATGGGCAATGCAATTGTA 201  
Db 151 ATTCGGGGTACCCCTATTGGCGCTGCGCGACGATGAATACAGCTTAAGCTTTACT 210  
Qy 202 CGCCATATACATGAGAACTCATTTGAGTGTGATGGCCCTCAAGCTAAGGCAATTCAGAT 261  
Db 211 TCACGCTATCAGAGAAATTCGTTGAGCTAGGTGACGACCTATATGCTGCTCCAGCTG 270  
Qy 262 GCAGCCAAAGCGGTGGGAATCATGTGTACCTGGGGATGATGAACGAGTCCGTGGACCC 321  
Db 271 GCGCGCGCGCGCAACAAATCGCACTCGTCATGGGCTATTCGAGCGGGAACCGGATCG 330  
Qy 322 CTTTACATCATGATGTTGTTTATAGCCGTAATGTTGACACATTTGGGGCCCGGGAAAG 381  
Db 331 CGCTATCTGAGCCAGGTGTTTATTCAGACGAGCTGGCGAGATGTTGCCAATGGCGCAAG 390  
Qy 382 TTGAACCTACTTTTGTGGAACGTACTTGTTCGGGGAAGGGAGTGTTCATCGCTAGCG 441  
Db 391 CTGAAGCCCAACAGCTTGAAGCTATGATCTACGCGGGAAGGCAACGAAACGATTTCTTC 450  
Qy 442 GTTTTCGAGACGCTGTGTGGAAGGCTGGGTGGCTTATGTGTTGGAGACCTTCAACCG 501  
Db 451 ACGCAACGACTTGGCGTTGCGAGCGGTGGGTGATGAACTGCTGGGAACATCTAACACCG 510  
Qy 502 CTAAACAAATACGCTTTGTATGACAAATGAAGATTCATTTGCGGCTTGGCGGAC 561  
Db 511 CTCACCAAGTTTATGATGATGACGCTCGGTGAGCAGGTTCACGTTGTCATCGTGGCGGCG 570  
Qy 562 TTTAGCCTTATCTAATGCGCGGAAGCGCTGGGGCTGATGTCAATGTAGCGGCTCT 621

Db 571 ATGTCCTCTTTCAGCCGGATGTTTTCCAACTGACATCGAAGCCAGCGGTACCC 630  
Qy 622 CGAATCTATGCGGTTGAAGGCAATGCTTCTGATAGGTGTGCGCTGTTTCAAA 681  
Db 631 GCGTGTGAGCAATCGAAGGCGCAACCTTTGTCTTGGCTGACGCAAGTATCGGACT 690  
Qy 682 TCCATGATCGAATGCTTTGTATAGATGACGAAAGATGCGGTTCTTGGCTGGTGT 741  
Db 691 AGCGGATCGAAGCTTGTGCTCAACGACGACGCGGCACTGTTCGCAAGAGAT 750  
Qy 742 GCACATCAGATCATGAGGCTGTGATGTGTGATCTTGTGCGGCTTTGGCAAAAT 801  
Db 751 GCGTGGCGCCCATTTAAGCCCGATGGAAGCGAGCTTGGAAAGCTCTGGCGGAAGAT 810  
Qy 802 GAAGGGTATTTCTTACGCAACCTTGTATCTGAGTAAGCATCTTGTCTTAATATGCG 861  
Db 811 GCTGAGGGGATCTTGTACGACGAGATCATGTGAGCAAGATTTGCTGCGCAAGGCTGGA 870  
Qy 862 GCAGACCTGTGCTGATTTATTCGCTCCGACATTAATCTGCTTGTATAGATGCGAC 921  
Db 871 GCGATCCGGTCCGCACTATTTCCGGGCTGACGTCGTGCTGCAATTCGACCCGCGC 930  
Qy 922 CCTAATTACCGGTAGTTGAATTGAAGGTGATCTTCG 959  
Db 931 AATCATACGCCAGTTCATCGCATCGCATTAAGCGTCG 968

## RESULT 4

US-10-919-182-3

; Sequence 3, Application US/10919182  
; Publication No. US20060035352A1  
; GENERAL INFORMATION:  
; APPLICANT: E.I. duPont de Nemours and Company, Inc.  
; APPLICANT: Di Cosimo, Robert  
; APPLICANT: Payne, Mark  
; APPLICANT: O'Keefe, Daniel  
; TITLE OF INVENTION: PRODUCTION OF 3-HYDROXYCARBOXYLIC ACID USING NITRILASE MUTANTS  
; FILE REFERENCE: CL/2584 US NA  
; CURRENT APPLICATION NUMBER: US/10/919,182  
; CURRENT FILING DATE: 2004-08-16  
; NUMBER OF SEQ ID NOS: 18  
; SOFTWARE: PatentIn version 3.2  
; SEQ ID NO: 3  
; LENGTH: 1110  
; TYPE: DNA  
; ORGANISM: Acidovorax facilis 72W  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: (1)..(1110)  
US-10-919-182-3

Query Match 21.5%; Score 218; DB 9; Length 1110;

Best Local Similarity 52.0%; Pred. No. 3.7e-60; Indels 0; Gaps 0;

Matches 488; Conservative 0; Mismatches 450; Indels 0; Gaps 0;

Qy 22 GCCTCGTGCAGAGCGCCCGATCTACATGATTTGGAGCGCGGTGAGCAAAACATT 81  
Db 31 GCAACCGTTGACGAGAGCGCGGTATGCTCGACGACGACGACATGACAAAGTCGATC 90  
Qy 82 GAGTTGATGAAGAAGACGACGATTAATATGCTGCTGATGCGCTTTCCGAAACTTGG 141  
Db 91 GGCATCATCGAAGAAGCTGCCAAAGGCGCGAGTGTATGCTTTCCGAAAGTATTC 150  
Qy 142 ATTCAGGTACCCATGTTTCTTTGGCTTGACTACACGACATGGGCAATGCAATTGTA 201  
Db 151 ATTCGGGGTACCCCTATTGGCGCTGCGCGACGATGAAGTACAGCTTAAGCTTTACT 210  
Qy 202 CGCCATATCAGTGAAGAACTCATTTGAGTGTGATGGCCCTCAAGCTAAGGCAATTCAGAT 261  
Db 211 TCACGCTATCAGAGAAATTCGTTGAGCTAGGTACGACGATATGCTGCTCCAGCTG 270  
Qy 262 GCAGCCAAAGCGGTGGGAATCATGTGTACCTGGGGATGATGAACGAGTCCGTGGACCC 321

Db 271 GCCGCGCGCCGACAAATGCACTGCTCATGGGCTATTTCGGAGCGGGAAGCCGGATCG 330  
 QY 322 CTTTACATCGTGTGAGTTCATAGGCGATATGATGACACCAATGGGGCCCGCGAAAG 381  
 Db 331 CGCTATCTGACCCAGGCTTCATGACGAGGTGGCGAAGATCGTTGCAATTCGGCGCAAG 390  
 QY 382 TTGAAACCTACTTTTGTGGAAGTACTTTGTCGGCGAAGGGAGTGGTTCAATCGCTAGCG 441  
 Db 391 CTGAAGCCCAACACAGCTGAGCGTACGATCTACGCGCAAGCAACGCAATTTCTC 450  
 QY 442 GTTTTCGAGAGCTGTGTGGAAGGCGTGGGCTTATGCTGTGTGGAGACCTTCAACCG 501  
 Db 451 ACGACGACTTGGCTGTGGAAGCGTGGTGGATTTGAATCTGCTGGAAACATTTCAACCG 510  
 QY 502 CTAAACAATAAGCTTTGTATGACCAAAATGAAGAATTCATTTGGCGCTTGGCCGAGC 561  
 Db 511 CTCAGCAAGTTCATGATGATGACGCTCGGTGAGCGAGTTCAGTTCATCTGGCCGCGG 570  
 QY 562 TTTAGCCTTTATCTTAATGCGCGCAAGCCCTGGGGCCCTGATGTCATGTAAGCGGCTCT 621  
 Db 571 ATGTCCCTCTTCAAGCCGAGTGTTCCTCAACTGAGCATCGAAGCAACGCGAGTCAAC 630  
 QY 622 CGAATCTATGCGGTTGAAGGGCAATGCTTGTACTAGCGTGTGGCTGCTTTCAACA 681  
 Db 631 CCTGTGACGAATGGAAGCCAAACCTTTGTGCTTGCACGCAAGGTGATCGGACT 690  
 QY 682 TCCATGATGATATGCTTTGTACAGATGACGAAAGCATGCGTTGCTTGGCTGTGGT 741  
 Db 691 AGGCGCATGGAACGTTGTGCTCAACGAGCAAGCGGCACTGTTCGCGAAGAGAT 750  
 QY 742 GGACACTACAGTATCATAGAGGCTGATGTGTGTGACTGTGGCGCTTGGCGAAAT 801  
 Db 751 GGTGGGCGCGCATTTACGCGCCGAGTGAAGGAGCTTGCAGAACCTTGTGCGAAAT 810  
 QY 802 GAAAGGGATTTCTTACGCAAACTTGAATCTGAGATGAGCATCTTGTAAATGGCG 861  
 Db 811 GCTGAGGGGATCTTGTACGCAAGATGATCTGAGACAGATTTCTGTGCGAAGCTGGA 870  
 QY 862 GCAGACCCGCTGTGCTATTTCCGTCGCAATTAATCTGTGTAATGATGCGACG 921  
 Db 871 GCGGATCCGCTGCGGACATTTGCGCGGCTGACGTGCTGTCAGTTGCAACCGCGC 930  
 QY 922 CTTAATTACCGGTAGTTGAATTTGAAGTGAATCTTCG 959  
 Db 931 AATCATACGCCAGTTTCATCGCATCGCATGACGCTG 968  
 RESULT 5  
 US-10-919-182-5  
 ; Sequence 5, Application US/10919182  
 ; Publication No. US2006003532A1  
 ; GENERAL INFORMATION:  
 ; APPLICANT: E.I. duPont de Nemours and Company, Inc.  
 ; APPLICANT: Di Cosimo, Robert  
 ; APPLICANT: Payne, Mark  
 ; APPLICANT: O'Keefe, Daniel  
 ; TITLE OF INVENTION: PRODUCTION OF 3-HYDROXYCARBOXYLIC ACID USING NITRILASE MUTANTS  
 ; FILE REFERENCE: C12584 US NA  
 ; CURRENT APPLICATION NUMBER: US/10/919,182  
 ; CURRENT FILING DATE: 2004-08-16  
 ; NUMBER OF SEQ ID NOS: 18  
 ; SOFTWARE: PatentIn version 3.2  
 ; SEQ ID NO 5  
 ; LENGTH: 1110  
 ; TYPE: DNA  
 ; ORGANISM: Artificial sequence  
 ; FEATURE:  
 ; OTHER INFORMATION: Mutant nitrilase B2 and H9  
 ; NAME/KEY: CDS  
 ; LOCATION: (1)..(1110)  
 US-10-919-182-5

Query Match 21.5%; Score 218; DB 9; Length 1110;  
 Best Local Similarity 52.0%; Pred. No. 3,7e-60;  
 Matches 488; Conservative 0; Mismatches 450; Indels 0; Gaps 0;  
 QY 22 GCTTCGTGCAAGCCGCCGATCTACATGATTTTGGAGCGGAGCTGACCAAAACCAT 81  
 Db 31 GCACCGTTCAGGCGAGACCCGATAGCTCGACCGACGACGACGACGATCGACATCGATC 90  
 QY 82 GAGTTGATGGAAGAAGACGACGTAATATGCTGCTGATGCTGCTTTCGGAAACTTGG 141  
 Db 91 GGCATCATGAAAGAGGCTGCCAAAAGCGGAGTCTGATGCTGCTTCCGAAATTC 150  
 QY 142 ATTCAGGCTACCATGCTTTTGGCTTGAATCAACAGATGGGCAATGCAATTTGTA 201  
 Db 151 ATTCGCGGCTACCCCTATTGGGCTGCTGCGAGCTGAAGTACAGCTTAAGCTTTACT 210  
 QY 202 CGCATACTACATGAGAACTCATTTGAGTGTGATGAGCCCTCAAGCTAAGCGCATTTCA 261  
 Db 211 TCACGCTATACAGAAATTCGTTGAGCTGAGTGAACGATGCTGCTCCAGCTG 270  
 QY 262 GCAGCAAGCGGTGGGAATCATGCTCACTGGGGAGTGAAGGAAGGGGTGGGCAAC 321  
 Db 271 GCCGCGCGCGCAACAAATTCGACTGTCATGAGGCTATTGGAAGCGGAGCCGATG 330  
 QY 322 CTTTACATCAGTACGATGTTTACAGGCGATTAATGTGACCACTTGGGCGCGGAAAG 381  
 Db 331 CGCTATCTGAGCGAGGCTTTCATCGACGAGCGTGGAGATGCTTGCATAGCGCGCAAG 390  
 QY 382 TTGAAACCTACTTTTGTGGAAGTACTTTGTCGGCGAAGGGAGTGGTTCAATCGCTAGCG 441  
 Db 391 CTGAAGCCCAACACAGCTGAGCGTACGATCTACGCGCAAGCAACGCAATTTCTC 450  
 QY 442 GTTTTCGAGAGCTGTGTGGAAGGCGTGGGCTTATGCTGTGTGGAGACCTTCAACCG 501  
 Db 451 ACGACGACTTGGCTGTGGAAGCGTGGTGGATTTGAATCTGAGCATCGAAGCAATTTCAACCG 510  
 QY 502 CTAACAATAATGAGCTTTGTATGACCAAAATGAAGAATTCATTTGGGCTTGGCCGAGC 561  
 Db 511 CTAAGCAAGTTCATGATGATGACGCTCGGTGACAGAGTTCATGCTGCGAAGCTGGA 570  
 QY 562 TTTAGCCTTTATCTTAATGCGCGCAAGCCCTGGGGCTGATGTCATGTAAGCGGCTCT 621  
 Db 571 ATGTCCCTCTTCAAGCCGAGTGTTCCTCAACTGAGCATCGAAGCAACGCGAGTCAAC 630  
 QY 622 CGAATCTATGCGGTTGAAGGGCAATGCTTGTACTAGCGTGTGGCTGCTTTCAACA 681  
 Db 631 CGCTGTGACGATCGAAGGCGCAACCTTGTGCTTGTGTCGACGAGGTGAGACT 690  
 QY 682 TCCATGATGATATGCTTTGTACAGATGAGGAAAGCATGCGTGTGGCTGTGGT 741  
 Db 691 AGCGGATCGAAAGCTTGTGCTCAACGACGACGCGCACATGTTCCGCAAGATGT 750  
 QY 742 GGACACTACGATATCAGGCGCTGATGTGTGATCTTGTGCGCGCTTGGCGAAAT 801  
 Db 751 GGTGGGCGCATTTACGCGCCGAGTGAAGCGAGCTTGGGAGCCCTGCGCGAAGAT 810  
 QY 802 GAAAGGGATTTCTTACGCAAACTTGAATCTGAGATGAGCATCTTGTAAATGGCG 861  
 Db 811 GCTGAGGGGATCTTGTACGCAAGATGATCTGAGACAGATTTCTGTGCGAAGCTGGA 870  
 QY 862 GCAGACCCGCTGTGCTATTTCCGTCGCAATTAATCTGTGTAATGATGCGACG 921  
 Db 871 GCGGATCCGCTGCGGACATTTGCGCGGCTGACGTGCTGTCAGTTGCAACCGCGC 930  
 QY 922 CTTAATTACCGGTAGTTGAATTTGAAGTGAATCTTCG 959  
 Db 931 AATCATACGCCAGTTTCATCGCATCGCATGACGCTG 968  
 RESULT 6  
 US-10-919-182-11  
 ; Sequence 11, Application US/10919182  
 ; Publication No. US2006003532A1

```
/ GENERAL INFORMATION:
/ APPLICANT: E.I. duPont de Nemours and Company, Inc.
/ APPLICANT: E.I. duPont, Robert
/ APPLICANT: Di Cosimo, Robert
/ APPLICANT: Payne, Mark
/ APPLICANT: O'Keefe, Daniel
/ TITLE OF INVENTION: PRODUCTION OF 3-HYDROXYCARBOXYLIC ACID USING NITRILASE MUTANTS
/ FILE REFERENCE: CL2584 US NA
/ CURRENT APPLICATION NUMBER: US/10/919,182
/ CURRENT FILING DATE: 2004-08-16
/ NUMBER OF SEQ ID NOS: 18
/ SOFTWARE: Patencin version 3.2
/ SEQ ID NO 11
/ LENGTH: 1110
/ TYPE: DNA
/ ORGANISM: Artificial sequence
/ FEATURE:
/ OTHER INFORMATION: Mutant nitrilase with Thr210 to Cys change
/ NAME/KEY: CDS
/ LOCATION: (1) .. (1110)
US-10-919-182-11

Query Match      21.5%; Score 218; DB 9; Length 1110;
Best Local Similarity 52.0%; Pred. No. 3.7e-60;
Matches 488; Conservative 0; Mismatches 450; Indels 0; Gaps 0;

QY 22 GCCTGCGTCAAGCCGCCGATCTACATGATTTGAGGCGGACGCTGACCAAAACATT 81
DB 31 GCAACCGTTCAGGCAAGCCGGTATGCTGACGAGAGCAACATCACAAGTCGATC 90
QY 82 GAGTTGATGGAAGAAGCAGCAGTAATAATGCTGCTGATGCGCTTCCGAAACTTGG 141
DB 91 GGCATCATGGAAGAAAGCTCCCAAAAAGGCGGAGTCTTAATGCTTCCGAAATTC 150
QY 142 ATTCAAGGCTACCATGTTCTTTGCTTGACTGACCAACGACATGGCAATTGTA 201
DB 151 ATTCGGGGCTACCCCTATGAGGCGTGGCTGGCGACGTGAAGTACAGCTTAAGCTTACT 210
QY 202 CGCCAAATCCATGAAACCTATTGAGTGGATGGCCCTCAAGCTAAGGCAATTCAGAT 261
DB 211 TCACGCTATCAGAGAAATTCGTTGAGCTAGGAGCAGACCGTATGCGTCCAGCTG 270
QY 262 GCAGCCAAAGCGGTTGGGATCATGCTGACCCCTGGGATGAGTGAACGGGTCGGTGCACC 321
DB 271 GCCGGCGCGCGCAAAATCGCATCTGCTATGGGCTATTCGAGCGGGAACCGGATCG 330
QY 332 CTTTACATCAGTCAGTGGTTTCATAGGCGCAATTAATGCTGACCACTTGGGCGCGAAAG 381
DB 331 CGCTATCTGAGCCAGGTGTTTCATGACGAGCGTGGGAGATGCTTGGCCAAATGGGGCAAG 390
QY 382 TTGAAACTACTTTTGTGTAAGTATCTTGTTCGGCGAAAGGAGATGTTCAATCGCTAGCG 441
DB 391 CTGAAGCCCAACAGCTGAGCGTACGATCTACGCGGAGGCAAGCAACCCATTTCCTC 450
QY 442 GTTTTCGAGAGCTGTTGGAAGGCGTGGGCTTATGCTGTTGGAGACCTTCAACCG 501
DB 451 AGCGACGACTTGGCTTCGAGCGCTCGGTGATTTGAATGCTGGAAATTTTCAACCG 510
QY 502 CTAAACAAATACGCTTTGTATGACCAAAATGAAGATTCATTGTGCGGCTTGGCGGAGC 561
DB 511 CTCACCAAGTTATCATATGTATACAGCGTCGAGTGAAGGTCACAGTTGCAATCGTGGCGGG 570
QY 562 TTTAGCTTTATCTTAATGCGGCGAAAGCCCTGGGAGCTGATGTCATATGTAAGCGGCTCT 621
DB 571 ATGTCCCTCTTTCACCCGATGTTTTCCAACTGACATCGAAGCCACGCGTCTGC 630
QY 632 GCAATCTATGCGGTTGAAGGCAATGCTTCTACTAGCGGTGAGCGCTGCTTTCAAA 681
DB 631 CGCTGTACGCAATGGAAGGCAAACTTTTGCTTTGCTTCAACGAGGTATCGGACT 690
QY 682 TCCATGATCGATATCTTTGTACAGATGACGAAAGCATGCGTTCCTTGGCTGTGT 741
DB 691 AGCGGATCGAAAGCTTCTGCTCAACGACGAAAGCGCGGCACTGTGCGGCAAGATGT 750
```

```
QY 742 GGACATCAAGTATCATAGGCGCTGATGAGTGAATTGTCGCGCTTCCGAAAT 801
DB 751 GCTGGGCGCGCATTTACGCGCCGATGAGAGCGAGCTTGCAGAGCTTGGCGAGAT 810
QY 802 GAAGAGGTATTTCTTACGCAAACTTGAATCTGAGTACGCAATCTTGTAAATGGCG 861
DB 811 GCTGAGGGGATCTTACGACAGATGATCTGAGCAGATTCCTGCGGCAAGGCTGGA 870
QY 862 GCAGACCTGCTGTGATATTTCCGTCGCCGACATTAATCTGCTTGAATAGATGACAC 921
DB 871 GCGGATCCGCTGCGGCACTATTCGCGGCTGACGTGCTGTGCTCAGTTGACCCGCGC 930
QY 922 CCTAATTAACGGGTAGTGAATGAGGTATCTTCG 959
DB 931 AATCATAGCCAGTTCATCGCATCGGCAATGACGCTG 968

RESULT 7
US-10-919-182-7
/ Sequence 7, Application US/10919182
/ Publication No. US20060035352A1
/ GENERAL INFORMATION:
/ APPLICANT: E.I. duPont de Nemours and Company, Inc.
/ APPLICANT: Di Cosimo, Robert
/ APPLICANT: Payne, Mark
/ APPLICANT: O'Keefe, Daniel
/ TITLE OF INVENTION: PRODUCTION OF 3-HYDROXYCARBOXYLIC ACID USING NITRILASE MUTANTS
/ FILE REFERENCE: CL2584 US NA
/ CURRENT APPLICATION NUMBER: US/10/919,182
/ CURRENT FILING DATE: 2004-08-16
/ NUMBER OF SEQ ID NOS: 18
/ SOFTWARE: Patencin version 3.2
/ SEQ ID NO 7
/ LENGTH: 1110
/ TYPE: DNA
/ ORGANISM: Artificial sequence
/ FEATURE:
/ OTHER INFORMATION: Mutant nitrilase B4
/ NAME/KEY: CDS
/ LOCATION: (1) .. (1110)
US-10-919-182-7

Query Match      21.3%; Score 216.4; DB 9; Length 1110;
Best Local Similarity 51.9%; Pred. No. 1.2e-59;
Matches 487; Conservative 0; Mismatches 451; Indels 0; Gaps 0;

QY 22 GCCTGCGTCAAGCCGCCGATCTACATGATTTGAGGCGGAGCGTGAACAAACATT 81
DB 31 GCAACCGTTCAGGCAAGCCGGTATGCTGACGAGCAACGATGACCAAGTCGATC 90
QY 82 GAGTTGATGGAAGAAGCAGCAGTAATAATGCTGCTGATGCGCTTCCGAAACTTGG 141
DB 91 GGCATCATGGAAGAAAGCTCCCAAAAAGGCGGAGTCTTAATGCTTCCGAAATTC 150
QY 142 ATTCAAGGCTACCATGTTCTTTGCTTGACTGACCAACGACATGGCAATTGTA 201
DB 151 ATTCGGGGCTACCCCTATGAGGCGTGGCTGGCGACGTGAAGTGAAGCTTAAGCTTACT 210
QY 202 CGCCAAATCCATGAAACCTATTGAGTGGATGGCCCTCAAGCTAAGGCAATTCAGAT 261
DB 211 TCACGCTATCAGAGAAATTCGTTGAGCTAGGTATGACCGCTATGCGTCCAGCTG 270
QY 262 GCAGCCAAAGCGGTTGGGATCATGCTGACCCCTGGGATGAGTGAACGGGTCGGTGCACC 321
DB 271 GCCGGCGCGCGCAAAATCGCACTGCTATGGGCTATTCGAGCGGGAAGCGGATCG 330
QY 332 CTTTACATCAGTCAGTGGTTTCATAGGCGCAATTAATGCTGACCACTTGGGCGCGAAAG 381
DB 331 CGCTATCTGAGCCAGGTGTTTCATGACGAGCGTGGCGAGATGTTGCCAAATGGGCGCAAG 390
QY 382 TTGAAACTACTTTTGTGTAAGTATCTTGTTCGGCGAAGGAGATGTTTATTCGCTACG 441
```

```
Db 391 CTAAGCCACACACGTTGAGCGTACGATCTACGCGCAAGCAACGGAATTCCTC 450
Qy 442 GTTTTGAGACGTCGTGTTGGAAGCTGGTGGCTTAATGCTTTGGAGACACTTCAACG 501
Db 451 ACGCAGCAGCTTCGGGTTCCGACGCGTGGATGGAATGCTGGGAACATTTCACACG 510
Qy 502 CTAACAATAATACGCTTTGATGACAAATAAGAAAGATTCTTTGGCGCTTGGCCGAGC 561
Db 511 CTCAGCAGATCATGATGTATACGCTCGGTGAGAGAGCTCCAGTTGATCGTGGCCGCG 570
Qy 562 TTTAGCCTTTATCTTAATGCGCGAAAGCCTTGGGCGCTGATGCAATGAGCGGCTCT 621
Db 571 ATGTCCCTCTTCAAGCCGAGATGTTTCCACTGAGCATGGAAGCCACGCGAGTATC 630
Qy 622 CGAATCTATGCGCTTGAAGGGCAATGCTTGTACTAGCGTGTGCGCTGCTTTCACAA 681
Db 631 CGCTCGTACGCAATGGAAGGCCAAACCTTTGTGCTTTCGACGAGGATATCGGACCT 690
Qy 682 TCCATGATCGATATGCTTTGTATACAGATGACGAAAGATGCGTTGCTTCTGGCTGTGT 741
Db 691 AGCGGATCGAAGCTTCTGCTCTCAAGACGAGCGCGACTGTGCTCCGCAAGATGT 750
Qy 742 GGAACATCACTATCATAGGCGCTTATGCTGTGCTGTGCGCGCTTTCGCGAAAT 801
Db 751 GGTGGGCGCGCATTTTACGCGCGGATGGAAGCGAGCTTGGCAAGCTCTGGCGAAGAT 810
Qy 802 GAAGAGGATATCTCTACGCAAACTTGATCTGAGTACGATCGCTTGTCTAAATGCGG 861
Db 811 GCTGAGGGGATCTTGTACCGACAGATCATCTGAGCAGATTCGTGCGGAGGCTGGA 870
Qy 862 GGAAGCCTGCTGTGCTATTTCCGCTCCGACATTACTCGTGTCTAATGATGCGAGC 921
Db 871 GCGATCCGCTGGGAGCACTATTCGCGCTGACGCTGTGCTGATCGATTCGACCCGCGC 930
Qy 922 CCTAATTACCGGTAGTTGAATGGAAGTGAATCTTG 959
Db 931 AATCATACGCGAGTTTCATCGCATTCGCAATTCGAGCTG 968
```

```
RESULT 8
US-10-919-182-13
; Sequence 13, Application US/10919182
; Publication No. US2006003532A1
; GENERAL INFORMATION:
; APPLICANT: E.I. duPont de Nemours and Company, Inc.
; APPLICANT: Di Cosimo, Robert
; APPLICANT: Payne, Mark
; APPLICANT: O'Keefe, Daniel
; TITLE OF INVENTION: PRODUCTION OF 3-HYDROXYCARBOXYLIC ACID USING NITRILASE MUTANTS
; FILE REFERENCE: CL2584 US NA
; CURRENT APPLICATION NUMBER: US/10/919,182
; CURRENT FILING DATE: 2004-08-16
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 13
; LENGTH: 1110
; TYPE: DNA
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Mutant nitrilase with Phe168 to Lys change
; NAME/KEY: CDS
; LOCATION: (1)..(1110)
US-10-919-182-13
```

```
Query Match 21.3%; Score 216.4; DB 9; Length 1110;
Best Local Similarity 51.9%; Pred. No. 1,2e-59;
Matches 487; Conservative 0; Mismatches 451; Indels 0; Gaps 0;
```

Qy 22 GCTGCGTGCAGAGCGCCCGATCTACATGATTTTGAAGGCGAGCGGGAACCAACTT 81  
Db 31 GCAACCGTTTCAGGAGAGCGCGGTATGGCTCGACGACAGCGAACGATCGACAACTCATC 90

```
Qy 82 GAGTGTATGGAAGAACACACGATTAATGCTGCTGTGATGCGCTTCCGGAACCTTG 141
Db 91 GGCATCATGAAAGAGAGCTGCCAAAGGCGGAGCTGATGCTTTCCCGAAGATTC 150
Qy 142 ATTCGAGGCTACCGATGTTTCTTTGGCTTGACTGACGACATGGGCAATGCAATTTGTA 201
Db 151 ATTCGAGGCTACCGCTTATTTGGGCGTGGCTTGGCGACGTGAAGTACAGCTTAAGCTTACT 210
Qy 202 CGCAATACATGAGAACTCATTTGATGATGAGTGGCCCTCAAGCTTAAGGCGATTCAGAT 261
Db 211 TCAGCTATCACGAGAAATTCGTTGAGCTAGGAGACGATGATGCTGTGCTCCAGCTG 270
Qy 262 GAGCGCAAGCGGTGGGAATCATGTCACCTTGGGAGATGAGGAAGGCTGGTGACAC 321
Db 271 GCGCGCGCGCAACAAATTCGACTGCTCATGAGGCTATTCGAGAGGGAAGCGGATG 330
Qy 322 CTTTACATCAGTACGTGTTTCATAGGCGATATATGATGACACATTCGGGCGCGCAAG 381
Db 331 CGCTATCTGAGCAGAGTTCATTCGACGAGCGTGGCGAGATCGTTCCAAATCGGCGCAAG 390
Qy 382 TTGAACCTTATCTTTTGAAGCTACTTGTGCGGGAAGGGAGTTCATGCTAGCG 441
Db 391 CTGAGGCCACACAGCTTGAAGCTATGATCTACGCGGAAGGCAACGAACTTCCTC 450
Qy 442 GTTTTGAGACGTCGTGGAAGGCTGGTGGCTTATGCTGTTGGAAGCACTTCAACG 501
Db 451 ACGCAGCATTCGCGTTCGGAACGCGTGGATGGAATGCACTGGGAACATTAACACG 510
Qy 502 CTAACAATAATACGCTTTGTATGCACAAAATGAAAGATTCATGCGCGCTTGGCGAGC 561
Db 511 CTCAGCAAGTTCATGATGTAAGCTCGGTGACAGATTCACGTTGATCTGAGCGCGG 570
Qy 562 TTTAGCCTTTATCTTAATGCGCGAAAGCCTTGGGCGCTGATGCAATGAGCGGCTCT 621
Db 571 ATGTCCCTCTTCAAGCCGAGATGTTTCCAACTGACATGGAAGCCACGCGAGTCAAC 630
Qy 622 CGAATCTATGCGCTTGAAGGCGAATGCTTGTACTAGCTGTGTGCGCTTTCACAA 681
Db 631 CGCTGTACGCAATCGAAGGCGAAACCTTTGTGCTTGTCTGACGAGGTATGACCT 690
Qy 682 TCCATGATCGATATGCTTTGTATACAGATGAGGAAAGATGCGTTCGTGGCTGTGT 741
Db 691 AGCGGATCGAAGGCTTCTGCTCAACGACGAGCGGCGACTGTGCCGCAAGATGT 750
Qy 742 GGAACATCACTATCATAGGCGCTTATGCTGTGATCTTGTGCGCTTTCGCGAAAT 801
Db 751 GGTGGGCGCGCATTTTACGCGCGGATGGAAGCGAGCTTGGCAAGCTCTGGCGAAGAT 810
Qy 802 GAAGAGGATATCTCTACGCAAACTTGATCTTGAAGTACGATCTTGTCTAAATGCGG 861
Db 811 GCTGAGGGGATCTTGTACGAGATGATTCGATTCGAGCAGATTCGTGCGAAGGCTGGA 870
Qy 862 GGAAGCCTGCTGTGCTATTTCCGCTCCGACATTACTCGTGTCTAATGATGCGAGC 921
Db 871 GCGATCCGCTGGGACATTAATTCGGGCTGACGTCGTGATCGAATTCGACCCGCGC 930
Qy 922 CCTAATTACCGGTAGTTGAATGGAAGTGAATCTTG 959
Db 931 AATCATACGCGAGTTTCATCGCATTCGCAATTCGAGCTG 968
```

```
RESULT 9
US-11-096-568A-6953
; Sequence 6953, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nikolai et al.
; TITLE OF INVENTION: Sequence-determined DNA Fragments and Corresponding Polypeptides
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
```

NUMBER OF SEQ ID NOS: 34471  
; SEQ ID NO 6953  
; LENGTH: 1415  
; TYPE: DNA  
; ORGANISM: glycine max  
; FEATURE:  
; NAME/KEY: misc.feature  
; LOCATION: (1)-(1415)  
; OTHER INFORMATION: Ceres Seq. ID no. 15169316  
US-11-096-568A-6953

Query Match 8.2%; Score 83; DB 11; Length 1415;  
Best Local Similarity 50.1%; Pred. No. 4.5e-16;  
Matches 271; Conservative 0; Mismatches 255; Indels 15; Gaps 2;

Qy 23 CCTGCTGCAAGCCGCCGATCTACATGATTTGAGCGGATGACAAACCAATTG 82  
Db 185 CCGTGTTCAGCCCTCCACATCTTCTACGACACCACTTAGATAGGCTGAGA 244  
Qy 83 AGTTGATGGAAGACAGACGTAATTAATGCTGCTGATGCGCTTTCGGAACCTTGA 142  
Db 245 GGTGTGTGCTGACATCTACATGATGAGTCCAGCTGTGTGTTTCCAGAGCTTTG 304  
Qy 143 TTCAGGCTACCCATGCTT---CTTGGCTTGAATCAACGACATGGGCAATGCAATT 198  
Db 305 TGGGTGGCTACCCGGTGGTTCAGCTTTTGTCTTTCGATGGGAATGGCACTGTTAAG 364  
Qy 199 GTA-----CGCAATACATGAGAACTCATTTGGATGGATGGCCCTCAAGCTA 247  
Db 365 GTAGAGAGAGTTTTCGCAAGTATCATTTCTGACGCAATGATGTGCTGGTCTCGAAGTTG 424  
Qy 248 AGCGATTTTCAGATGACGCAAGCGGTTGGGAATCATGCTCACCTGGGGAATGATGAC 307  
Db 425 ATAGTTGGCGACCAATGGAGGAAATTAAGTACATTTAGTGAATGGGTGATGATGAGA 484  
Qy 308 GGGTGTGTCGACCTTTTACATCACTGATGATGATGATGATGATGATGATGATGATG 367  
Db 485 GGGAGGCTACACATTTATTTGACCGCTTCTTCTTGTGATTTCTAGGATCTTACCTAG 544  
Qy 368 GGGCCCGGCAAGTTGAAACCTACTTTTGTGAAAGTACTTTTGTGGCGAAGGGATG 427  
Db 545 GAAAGCAGAGAAATCATGCTGATGATGACGCGGTTATCTGGGGATTTGGGGATG 604  
Qy 428 GTTATGCTTACCGCTTTTCCAGAGTCTGTTGGAAGGCTGGTGGCTTATGCTTGGG 487  
Db 605 GATCAACCATTCATGATGTTTGAACCTCCGTTGAAAAATAGTGTCTCCATTGTTGGG 664  
Qy 488 AGCACTTCAACCGCTTACAAATACTGTTGTATGCACAAAATGAGAGATTCATTGG 547  
Db 665 AGAACGAAATGCACTATTAAAGACAGCAATGATGCAAAAGTGTGAGATATATTGTG 724  
Qy 548 C 548  
Db 725 C 725

RESULT 10  
US-11-096-568A-20685  
; Sequence 20685, Application US/11096568A  
; Publication No. US20060048240A1  
; GENERAL INFORMATION:  
; APPLICANT: Alexandrov, Nickolai et al.  
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides  
; FILE REFERENCE: 2750-1592PUS2  
; CURRENT APPLICATION NUMBER: US/11/096,568A  
; CURRENT FILING DATE: 2005-04-01  
; NUMBER OF SEQ ID NOS: 34471  
; SEQ ID NO 20685  
; LENGTH: 1706  
; TYPE: DNA  
; ORGANISM: Zea mays subsp. mays  
; FEATURE:

NAME/KEY: misc.feature  
; LOCATION: (1)-(1706)  
; OTHER INFORMATION: Ceres Seq. ID no. 12386684  
US-11-096-568A-20685

Query Match 8.0%; Score 81.6; DB 11; Length 1706;  
Best Local Similarity 49.8%; Pred. No. 1.4e-15;  
Matches 272; Conservative 0; Mismatches 259; Indels 15; Gaps 2;

Qy 18 GGTGCTGCTGCAAGCCGCCGATCTACATGATTTGAGCGGATGACAAAC 77  
Db 629 GGTCAACCGCTGTCAGCGGCTGCTGCTGTTTACGACACCCCTGGACCTCTGATTAAGC 688  
Qy 78 CATGATGATGGAAGAGACAGCACTTAATGCTGCTGATGCTGCTTTCGGGAAC 137  
Db 689 GAGAAATGATGAGAGAGAGCTGGGTATGTTTCAAGATGTTGTTTTCGGAAAGT 748  
Qy 138 TTGATTCAGAGCTACCA-----TGGTTCTTGTGCTTGAATCAAC-CAGC 182  
Db 749 CTTTGTGTGCTGCTACCAACATGATCTACCTTTGATGATGATGATGATGATGATGATG 808  
Qy 183 ATGGCAATGCAATTTGTACGCCAATACATGAGAACTCATTTGAGATGAGCCCTCA 242  
Db 809 CAAGGAAAGGAAGACTTTAGAAATGATCAGCATCTGCTAGATGCTGCTGCTCAGA 868  
Qy 243 AGCTAAGGCAATTCAGATGAGCCAGCGGTTGGGAATCATGCTCACCTGGGGAATGAG 302  
Db 869 AGTGTCCGCTTATCTGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 928  
Qy 303 TGAAGGCTGCTGTCACCTTTACATCACTGATGATGATGATGATGATGATGATGATGATG 362  
Db 929 TGAAGGCGAGATATACATCTTACACACGCTGCTCTCTGATCCTGAGGAAATTA 988  
Qy 363 CATTGGGCGCCGCAAGTTGAAACCTACTTTTGTGAAAGTACTTTTGTGGCGGAAG 422  
Db 989 CCTAGGGAAGACCGCAAGGCTATGCTATGCACTGACATGAAAGTATCTGGGGGTTGG 1048  
Qy 423 GGATGCTTACGCTGATGCGGTTTTCAGAGCTCTGTTGGAAGGCTGGTGGCTTATGCTG 482  
Db 1049 AGATGATCTACATTAATCTGCTGATGATGATGATGATGATGATGATGATGATGATG 1108  
Qy 483 TTGGAGAGCACTTCAACCGCTTACAAATACTGTTGTATGACCAAAATGAGAGATTCA 542  
Db 1109 CTGGGAAACGAATATGCACTTCTCAGAGCGGCAATGATGCAAAAGTATGAGATATA 1168  
Qy 543 TTGTGC 548  
Db 1169 TTGTGC 1174

RESULT 11  
US-10-932-182A-166788  
; Sequence 166788, Application US/10932182A  
; Publication No. US20060046253A1  
; GENERAL INFORMATION:  
; APPLICANT: NAKAO, YOSHIHIRO  
; APPLICANT: NAKAMURA, NORIHIISA  
; APPLICANT: KODAMA, YUKIKO  
; APPLICANT: FUDIMURA, TOMOKO  
; APPLICANT: ASHIKARI, TOSHIHIKO  
; TITLE OF INVENTION: METHODS FOR ANALYZING GENES OF INDUSTRIAL YEASTS  
; FILE REFERENCE: 030685-043  
; CURRENT APPLICATION NUMBER: US/10/932,182A  
; CURRENT FILING DATE: 2004-09-02  
; NUMBER OF SEQ ID NOS: 197023  
; SOFTWARE: PatentIn version 3.3  
; SEQ ID NO 166788  
; LENGTH: 597  
; TYPE: DNA  
; ORGANISM: Saccharomyces pastorianus  
US-10-932-182A-166788  
Query Match 6.9%; Score 70.4; DB 9; Length 597;



```
QY      473 GCTTATGCTGTGGAGACCTTCAACCGCTTACAAATAAGCTTTGATGCACAAATG 532
      |||||
Db      417 GATCATGCTACATGTTCTATCTTAACAGAAACAAATGATTTTCTGGCTAAGGCA 358
      |||||

QY      533 AAGAGATTCAATTGCGGCTTGGCCGAGCTTTAGCCTTTATCTTAATGCGGCGA 586
      |||||
Db      357 AAATAATTCCTTGCTCTACTGTGTCCTCAATTTTGGCTATATCAAACTTAGTCCA 304
      |||||
```

## RESULT 15

```
US-10-301-480-489364/c
; Sequence 489364, Application US/10301480
; Publication No. US20060057564A1
; GENERAL INFORMATION:
; APPLICANT: Wang, David G.
; TITLE OF INVENTION: Identification and Mapping of Single Nucleotide Polymorphisms
; FILE REFERENCE: 108827.137
; CURRENT APPLICATION NUMBER: US/10/301,480
; PRIOR FILING DATE: 2002-11-21
; PRIOR APPLICATION NUMBER: US 10/215,598
; PRIOR FILING DATE: 2002-08-09
; PRIOR APPLICATION NUMBER: US 60/311,695
; PRIOR FILING DATE: 2001-08-10
; NUMBER OF SEQ ID NOS: 1226818
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 489364
; LENGTH: 563
; TYPE: DNA
; ORGANISM: Homo sapien
US-10-301-480-489364
```

```
Query Match      3.4%; Score 34; DB 10; Length 563;
Best Local Similarity 56.1%; Pred.No. 2.9;
Matches 64; Conservative 0; Mismatches 50; Indels 0; Gaps 0;
```

```
QY      473 GCTTATGCTGTGGAGACCTTCAACCGCTTACAAATAAGCTTTGATGCACAAATG 532
      |||||
Db      417 GATCATGCTACATGTTCTATCTTAACAGAAACAAATGATTTTCTGGCTAAGGCA 358
      |||||

QY      533 AAGAGATTCAATTGCGGCTTGGCCGAGCTTTAGCCTTTATCTTAATGCGGCGA 586
      |||||
Db      357 AAATAATTCCTTGCTCTACTGTGTCCTCAATTTTGGCTATATCAAACTTAGTCCA 304
      |||||
```

Search completed: April 27, 2006, 00:22:15  
Job time : 612.88 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: April 27, 2006, 00:55:21 ; Search time 25.164 Seconds  
(without alignments)  
1288.550 Million cell updates/sec

Title: US-09-751-299-4  
Perfect score: 1765  
Sequence: 1 MKEAIKVCACVQAPITMDL.....DLRPYALGXASETGAQLEEI 337

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283416 seqs, 96216763 residues

Total number of hits satisfying chosen parameters: 283416

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database :  
1: PIR.80:\*  
2: PIR1:\*  
3: PIR3:\*  
4: PIR4:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	836.5	47.4	366	2 A45070	nitrilase - Rhodoc
2	819	46.4	356	2 A47181	nitrilase (EC 3.5.
3	800	45.3	383	2 A43470	aliphatic nitrilase
4	638	36.1	354	2 JC4212	nitrilase (EC 3.5.
5	634.5	35.9	349	1 A28658	nitrilase (EC 3.5.
6	559	31.7	368	2 QJ1613	cyanide hydratase
7	523.5	29.7	346	2 S77025	nitrilase (EC 3.5.
8	499.5	28.3	346	2 T49147	nitrilase (EC 3.5.
9	493.5	28.0	346	2 S22398	nitrilase (EC 3.5.
10	492.5	27.9	339	2 S31969	nitrilase (EC 3.5.
11	491.5	27.8	339	2 T52259	nitrilase (EC 3.5.
12	490.5	27.8	346	2 T49148	nitrilase (EC 3.5.
13	489.5	27.7	339	2 T52262	nitrilase (EC 3.5.
14	453.5	25.7	305	2 T27679	probable nitrilase
15	451.5	25.6	362	2 T52266	nitrilase-like pro
16	445	25.2	348	2 T03739	nitrilase (EC 3.5.
17	443.5	25.1	355	2 T52265	nitrilase (EC 3.5.
18	427.5	24.2	349	2 T03736	nitrilase (EC 3.5.
19	257.5	14.6	199	2 S50363	nitrilase homolog
20	207	11.7	272	2 T41662	probable nitrilase
21	185	10.5	297	2 F75263	probable hydrolase
22	183	10.4	262	2 C71109	hypothetical prote
23	177	10.0	292	2 C87275	hydrolase, carbon-
24	172.5	9.8	318	2 T48563	hypothetical prote
25	171	9.7	262	2 C75051	hypothetical related
26	160	9.1	257	2 C69264	conserved hypochet
27	156.5	8.9	257	2 AH3225	amidohydrolase (im
28	152.5	8.6	285	2 S58240	hypothetical prote
29	152.5	8.6	295	2 H82556	beta-alanine synth

30	151.5	8.6	322	2 T38399	probable amidohydr
31	150.5	8.5	576	2 G72277	NH(3)-dependent NA
32	146.5	8.3	267	2 B72408	conserved hypochet
33	143	8.1	291	2 S51459	hypothetical prote
34	143	8.1	294	2 AB0115	probable carbon-ni
35	142.5	8.1	292	2 B64558	conserved hypochet
36	142	8.0	312	2 JW0083	N-carbamyl-D-amino
37	140.5	8.0	294	2 G71949	hypothetical prote
38	138.5	7.8	272	2 B69109	N-carbamyl-D-amino
39	138	7.8	280	2 T34905	probable hydrolase
40	136.5	7.7	393	2 S27881	beta-alanine synth
41	136	7.7	298	2 T17568	hydrolase homolog
42	135.5	7.7	290	2 B81369	probable hydrolase
43	134	7.6	220	2 A84673	probable nitrilase
44	134	7.6	276	2 T40601	putative nitrilase
45	132.5	7.5	292	2 G83608	probable hydratase

ALIGNMENTS

```
RESULT 1
A45070
nitrilase - Rhodococcus rhodochrous
C:Species: Rhodococcus rhodochrous
C>Date: 10-Jun-1993 #sequence_revision 18-Nov-1994 #text_change 31-Dec-2004
C:Accession: A45070
R: Kobayashi, M.; Komeda, H.; Yanaka, N.; Nagasawa, T.; Yamada, H.
J. Biol. Chem. 267, 20746-20751, 1992
A:Title: Nitrilase from Rhodococcus rhodochrous J1. Sequencing and overexpression of the
A:Reference number: A45070; MUID:93015976; PMID:11400390
A:Contents: J1
A:Accession: A45070
A:Status: preliminary
A:Molecule type: DNA, protein
A:Residues: 1-366 <KOB>
A:Cross-references: UNIPROT:Q03217; UNIPARC:UP1000016PF36; GB:D11425; NID:G216933; PIDD:116040
A>Note: Sequence extracted from NCBI backbone (NCBI:116039, NCBI:116040)
C:Superfamily: nitrilase (carbon-nitrogen hydrolase)

Query Match      47.4%; Score 836.5; DB 2; Length 366;
Best Local Similarity 50.6%; Pred. No. 2.9e-68;
Matches 156; Conservative 56; Mismatches 95; Indels 1; Gaps 1;

QY 6 KVACVQAPITMDLEATVDTITLMEBAANNAKRLAFPETWIPGYWFLMDSPPAWAM- 64
DB 9 KVAAVQAPWPFDAKTVDSIIAEAAANGCELVAFPVFIPGYHYIWDSPLAGMA 68
QY 65 QFVRQYHNSLELDGQAKRISDAKRLGIMVTLGMSERVSGTLYISQWFIQDNGDTIGA 124
DB 69 KFAVRHNSLTWDSPPVQRLDPAARDHNAIVVVGISERDGSGLYMTQVLIDAGQOLVAR 128
QY 125 RRLKLPFTFVERTLFGSDGSLAVFETSVGLRGLCWEHLQPLTKYALYAONEBHCIAA 184
DB 129 RRLKLPFTFVRSYVYEGNSDISVYDMPARLALNCWEHFPQTLTKYAMSMEHQVAVAS 188
QY 185 WPSFSLYPNAKALGPDVNVAAASRIYAVEGQCVLASCALVQSQMTDMLCTDEKHALLL 244
DB 189 WPGMSLYQPEVPFAFGVDAQLTATRMVALBSQTFVVCCTGVVTPDEAHEFFCDNDEQRXLIG 248
QY 245 AGGSHRIITGPDGDLVAPLAENBEGILVYANLDPGVRIILAKMAADPAGHSRPDITLLT 304
DB 249 RGGGFARITIGPDGDLATPLADEBEGILVADIDLSAITLAKQADVPAGHSRDPVLSLNF 308
QY 305 DRSPKLPV 312
DB 309 NQRRHTPV 316

RESULT 2
A47181
nitrilase (EC 3.5.5.1), arylacetone-specific - Alcaligenes faecalis
C:Species: Alcaligenes faecalis
```

C>Date: 21-Sep-1993 #sequence\_revision 18-Nov-1994 #text\_change 31-Dec-2004  
C/Accession: A47181, S13860  
R/Kobayashi, M.; Izui, H.; Nagasawa, T.; Yamada, H.  
Proc. Natl. Acad. Sci. U.S.A. 90, 247-251, 1993  
A/Title: Nitrlase in biosynthesis of the plant hormone indole-3-acetic acid from indole  
A/Reference number: A47181, MUID:93126352; PMID:8419930  
A/Contents: JMJ  
A/Accession: A47181  
A/Status: preliminary  
A/Molecule type: DNA  
A/Residues: 1-356 <KOB>  
A/Cross-references: UNIPROT:P20960; UNIPARC:UPI000003514; GB:D13419; NID:9216202; PIDD:  
A/Note: sequence extracted from NCBI backbone (NCBIN:122081, NCBI:P.122082)  
R/Nagasawa, T.; Manger, J.; Yamada, H.  
Eur. J. Biochem. 194, 765-772, 1990  
A/Title: A novel nitrlase, arylacetone nitrlase, of Alcaligenes faecalis JMJ3. Purification  
A/Reference number: S13860; MUID:91099356; PMID:2269298  
A/Accession: S13860  
A/Status: preliminary  
A/Molecule type: protein  
A/Residues: 1-33 <NAG>  
A/Cross-references: UNIPARC:UPI0000175E31  
C/Superfamily: nitrlase (carbon-nitrogen hydrolase)  
C/Keywords: hydrolase

Query Match 46.4%; Score 819; DB 2; Length 356;  
Best Local Similarity 49.7%; Pred. No. 1,1e-66;  
Matches 156; Conservative 54; Mismatches 104; Indels 0; Gaps 0;

QY 2 KEATKACVQAAPRYMDLEATVTKTILMEBAARNARLAFPEWTWPGYPWFLMLDSPA 61  
DB 4 RKIVRAAVALAASPNYLATGVDTTELARQARDECDLVFGSTWLPFPFHVWLAGAPA 63  
QY 62 WAMOFVROYHENSLELDGPOAKRISDAKRLGIWVTLGMSERVGTLTYSQWPIGNDGT 121  
DB 64 WSLKYSARYANSLSDSAERQRIQAQARTLGIPTALGYSRSGSLYLGGCLIDDKGOM 123  
QY 122 IGARRKLTPTVERTLFGEGDSSLAIVETSVGLGICCEHIOPLTKYALYQNEEIH 181  
DB 124 LWSRRKLTPTVERTVGEGRDLVSDLELGVGLCCEHISPLSKALYSQHEAIIH 183  
QY 182 CAAMPSESLYNAKALGPDVNVAASRIYAVEGOCFYLASCAVSGMIDMLCTDDEKHA 241  
DB 184 IAMPSSISVSEQHHLISAKNMMAASQIYSVEGOCFYIASSSVTQGTIDMLLEVGEHNAS 243  
QY 242 LLLAGGHSRIIGPDGDLVAPLAENEBGIIYANLDPEVRIILAMADDPAGHSRPTDR 301  
DB 244 LLKVGGSSTMFAPDGRTLAPFLPHDAEGILIIADLNMEIAPAKAINDPVGHYSKPEATR 303  
QY 302 LLIDRSPKLPVVEI 315  
DB 304 LVLDLGHREPWTRV 317

RESULT 3  
A43470  
aliphatic nitrlase - Rhodococcus rhodochrous

C/Species: Rhodococcus rhodochrous  
C/Date: 10-Jun-1993 #sequence\_revision 18-Nov-1994 #text\_change 31-Dec-2004  
C/Accession: A43470  
R/Kobayashi, M.; Yanaka, N.; Nagasawa, T.; Yamada, H.  
Biochemistry 31, 9000-9007, 1992  
A/Title: Primary structure of an aliphatic nitrlase-degrading enzyme, aliphatic nitrlase  
A/Reference number: A43470; MUID:93003039; PMID:1390687  
A/Contents: K22  
A/Accession: A43470  
A/Status: preliminary  
A/Molecule type: DNA; protein  
A/Residues: 1-383 <KOB>  
A/Cross-references: UNIPROT:002068; UNIPARC:UPI0000130516; GB:D12583; NID:9216931; PIDD:  
A/Note: sequence extracted from NCBI backbone (NCBIN:114184, NCBI:P.114185)  
C/Superfamily: nitrlase (carbon-nitrogen hydrolase)

Query Match 45.3%; Score 800; DB 2; Length 383;  
Best Local Similarity 49.8%; Pred. No. 6,6e-65;  
Matches 159; Conservative 53; Mismatches 103; Indels 4; Gaps 4;

QY 5 IKVACVQAAPRYMDLEATVTKTILMEBAARNARLAFPEWTWPGYPWFLMLDSPA 64  
DB 13 VKVAVQAAPRYMDLEATVTKTILMEBAARNARLAFPEWTWPGYPWFLMLDSPA 72  
QY 65 -QFROYHENSLELDGPOAKRISDAKRLGIWVTLGMSERVGTLTYSQWPIGNDGTIG 123  
DB 73 SDIPIKYHENSLELDGPOAKRISDAKRLGIWVTLGMSERVGTLTYSQWPIGNDGTIG 132  
QY 124 ARRLKPTVERTLFGEGDSSLAIVETSVGLGICCEHIOPLTKYALYQNEEIHCA 183  
DB 133 NRKRLKPTVERTVGEGRDLVSDLELGVGLCCEHISPLSKALYSQHEAIIH 192  
QY 184 AMPSPSLYNAKALGPDVNVAASRIYAVEGOCFYLASCAVSGMIDMLCTDDE-KHA 241  
DB 193 SMPAMFALTPDVHQ-LSVEANDVTNYSVALIEGOTFLASTHVIKATQDLFAGDDAKGA 251  
QY 242 LLLAGGHSRIIGPDGDLVAPLAENEBGIIYANLDPEVRIILAMADDPAGHSRPTDR 301  
DB 252 LLPLGGGMARIYGPDKSLAEPLEDAEGILYALDEQIILAKAADDPAGHSRBDVLS 311  
QY 302 LLIDRSPKLPVVEI 320  
DB 312 LKIDTNNHTFVQYITADGR 330

RESULT 4  
J04212  
nitrlase (EC 3.5.5.1) - Comamonas testosteroni

C/Species: Comamonas testosteroni  
C/Date: 14-Nov-1995 #sequence\_revision 08-Feb-1996 #text\_change 31-Dec-2004  
C/Accession: J04212; PC4056  
R/Levy-Schli, S.; Soudrier, F.; Crutz-Le Coq, A.M.; Faucher, D.; Crouzet, J.; Petre, D.  
Gene 161, 15-20, 1995  
A/Title: Aliphatic nitrlase from a soil-isolated Comamonas testosteroni sp.: Gene cloning  
A/Reference number: J04212; MUID:95369726; PMID:7642130  
A/Accession: J04212  
A/Molecule type: DNA  
A/Residues: 1-354 <LEV>  
A/Cross-references: UNIPROT:Q59329; UNIPARC:UPI000008988B; GB:I32589; NID:91082008; PIDD:  
A/Accession: PC4056  
A/Molecule type: protein  
A/Residues: 154-162,295-302,323-340 <LE2>  
A/Cross-references: UNIPARC:UPI0000175E33; UNIPARC:UPI0000175E34  
C/Comment: This enzyme is active on adiponitrile and cyanovaletric acid.  
C/Genetics:  
A/Gene: nltA  
C/Superfamily: nitrlase (carbon-nitrogen hydrolase)  
C/Keywords: hydrolase  
F163/Active site: Cys #status predicted

Query Match 36.1%; Score 638; DB 2; Length 354;  
Best Local Similarity 41.2%; Pred. No. 3,4e-50;  
Matches 134; Conservative 62; Mismatches 107; Indels 22; Gaps 5;

QY 5 IKVACVQAAPRYMDLEATVTKTILMEBAARNARLAFPEWTWPGYPWFLMLDSPA 59  
DB 7 VKVAVQAAPRYMDLEATVTKTILMEBAARNARLAFPEWTWPGYPWFLMLDSPA 66  
QY 60 PAMAFVROYHENSLELDGPOAKRISDAKRLGIWVTLGMSERVGTLTYSQWPIGNDG 119  
DB 67 MMMAVLF-----KVAIEIPSEVQIISDAKKNGVYCVSSEKDNASLYLTQLMPFNG 121  
QY 120 DTIGARRKLTPTVERTLFGEGDSSLAIVETSVGLGICCEHIOPLTKYALYQNEE 179  
DB 122 NLIKHRRKPTPTSESRVWVGDDGSMAPVFTETEGNLGGIQCWEHALPLNIAAGSLNEQ 181  
QY 180 IHCAAMPSESLYNAKALGPDVNVA-----SRIYAVEGOCFYLASCAVSGSMI 230  
DB 182 VHVASWAPF--VPRGAVSSRVSSVCASTANMHQIIISQFVAINQVYVIMSTNLVGGQMI 239







A:introns: 44/1; 104/1; 202/1; 296/1  
C:Superfamily: nitrilase (carbon-nitrogen hydrolase)  
C:Keywords: hydrolase

Query Match 27.8%; Score 490.5; DB 2; Length 346;  
Best Local Similarity 35.7%; Pred. No. 9e-37;  
Matches 115; Conservative 65; Mismatches 113; Indels 29; Gaps 8;

QY 5 IKVACVQAPRYMDLEATVDKTIEMEDANNAARLLAFETWIPGYWFMIDSPAMM 64  
DB 25 VRVTIVQSTVYNDPATLIDRAEKRIYEAASKGALVFPAPFGTGRGFRGLAVGH 84  
QY 59 SPAMMQVROYHENSELDGPOAKRISDAKRLGIWTLGMSERVGTLIYISQWFI 118  
DB 85 NEEGRDER-KNYHASAIKVPGEVERKLAELAGKNVHLVWGALIKDGYTLCTALFPSPQ 143  
QY 119 GDTIGARRKLTPEVETLTFEGDSSLAFTSVGRIGLCCMEHLOPLTKVLYAONE 178  
DB 144 GGFLEKHKRVMPSTLERCIWGGDSTIPVYDTPIGKIGAAICWENRMPRLTALYAKGI 203  
QY 179 EIHCAAMPSEFSLYPNAAKALGPDVVAASRIY-AVEGQCFYLASCALVSQSMI----DML 233  
DB 204 EIVCAPTADYSL-----EMQSMHIAVEGGCFYLSAHQPKRREPENPDYL 251  
QY 234 ----CTDEKHALLLAGGHSRIIGPDGDLVAPLAENEGIIYANLDPCVRILAKMADP 290  
DB 252 FNDIVDTKHDPTVSGG-SVLIISPLGKVLGPNYES-EGLVTDLDLDIARAKLYPDV 309  
QY 291 AGHYSPDITRLILDRSPKLPV 312  
DB 310 VGHYSKPDIFNLTVNEHPKPV 331

RESULT 13  
T52262  
nitrilase (EC 3.5.5.1) 2 [imported] - Arabidopsis thaliana  
C:Species: Arabidopsis thaliana (mouse-ear cress)  
C>Date: 20-Oct-2000 #sequence\_revision 20-Oct-2000 #text\_change 31-Dec-2004  
C:Accession: T52262  
R:Bartel, B.; Fink, G.R.  
Proc. Natl. Acad. Sci. U.S.A. 91, 6649-53, 1994  
A:Title: Differential regulation of an auxin-producing nitrilase gene family in Arabidop  
A:Reference number: Z24515; MUID:8022831; PMID:8022831  
A:Accession: T52262  
A:Status: preliminary; translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-339 <BAR>  
A:Cross-references: UNIPROT:P32962; UNIPARC:UPI000016DA4C; EMBL:U38845; PIDD:AAE05220.1  
A:Experimental source: cultivar Columbia  
C:Genetics:  
A:Gene: NIT2  
A:Map position: 3  
A:introns: 37/2; 97/1; 195/1; 289/1  
C:Superfamily: nitrilase (carbon-nitrogen hydrolase)  
C:Keywords: hydrolase

Query Match 27.7%; Score 489.5; DB 2; Length 339;  
Best Local Similarity 33.1%; Pred. No. 1.e-36;  
Matches 112; Conservative 60; Mismatches 111; Indels 55; Gaps 7;

QY 5 IKVACVQAPRYMDLEATVDKTIEMEDANNAARLLAFETWIPGYWFMIDSPAMM 64  
DB 18 VRATIVQSTVYNDPATLIDRAEKRIYEAASKGALVFPAPFGTGRGFRGLAVGH 84  
QY 65 QF-----VQYHENSELDGPOAKRISDAKRLGIWTLGMSERVGTLIYI 110  
DB 69 RRGGLGVNNEGRDEFRKYHASAIKVPGEVERKLAELAGKNVHLVWGALIKDGYTLCT 128  
QY 111 SCWFIDNGDTIGARRKLTPEVETLTFEGDSSLAFTSVGRIGLCCMEHLOPLTK 170  
DB 129 TALFPSPQGFLEKHKRVMPSTLERCIWGGDSTIPVYDTPIGKIGAAICWENRMPYR 188  
QY 171 YALYQNEBHICA-----AMPSFSLYNAAKALGPDVVAASRIYAVEGQCFYLASC 223

DB 189 TALYAKGIELCAPTADSGKESQSMH-----IALEGCFVLASQ 230

QY 224 LVQSQSMI-----DMLCTD-DEKHALLAGGHSRIIGPDGDLVAPLAENEGIIYANLD 277  
DB 221 FCLRKDFPDHDDYLFTHWYDCKEPPDSIVSQGSVILSPLOVLGAPNYES-EGLTIRADD 289

QY 278 PGVRLAKMADPAGHYSPDITRLILDRSPKLPVVEI 315  
DB 290 LGDVARAKLYPDSVGHISRPVLAHLTVNEHPKRVTFI 327

RESULT 14  
T27679  
probable nitrilase (EC 3.5.5.1) ZK1058.6 - Caenorhabditis elegans  
C:Species: Caenorhabditis elegans  
C>Date: 15-Oct-1999 #sequence\_revision 15-Oct-1999 #text\_change 31-Dec-2004  
C:Accession: T27679  
R:Mortimore, B.  
Submitted to the EMBL Data Library, August 1994  
A:Reference number: Z20403  
A:Accession: T27679  
A:Status: translated from GB/EMBL/DBJ  
A:Molecule type: DNA  
A:Residues: 1-305 <MIL>  
A:Cross-references: UNIPROT:Q23384; UNIPARC:UPI000007C5FC; EMBL:Z35604; PIDD:CAA84681.1;  
A:Experimental source: clone ZK1058  
C:Genetics:  
A:Gene: CESP.ZK1058.6  
A:Map position: 3  
A:introns: 71/2; 178/1; 205/2  
C:Superfamily: nitrilase (carbon-nitrogen hydrolase)  
C:Keywords: hydrolase

Query Match 25.7%; Score 453.5; DB 2; Length 305;  
Best Local Similarity 35.5%; Pred. No. 1.8e-33;  
Matches 113; Conservative 61; Mismatches 119; Indels 25; Gaps 9;

QY 6 KVACVQAPRYMDLEATVDKTIEMEDANNAARLLAFETWIPGYW----FLMIDS 59  
DB 3 KIAIVQAGTLPFKPATLEKVKKNVEBAAGALVFPBAPFGIGPKNNNSGITWGTRT 62  
QY 60 PAMMQVROYHENSELDGPOAKRISDAKRLGIWTLGMSERVGTLIYISQWFI 119  
DB 63 PEGKKEF-KRIFENAIENGEBSLTLESIAQNNHIVIGVVERASTLYCSVFYSPSG 121  
QY 120 DTIGARRKLTPEVETLTFEGDSSLAFTSVGRIGLCCMEHLOPLTKVLYAONE 179  
DB 122 YLGGHRLTPTALERCVWGGDSTMPVSTSVKIGSAICWENRMPRLTALYAKGI 180  
QY 180 IHCAAMPSEFSLYPNAAKALGPDVVAASRIYAVEGQCFVLASQALVSQSMIDM---LCTD 236  
DB 181 IYLA--PTVD-----DRDVLSTMTRTALEGRFCFVSACQFLKSDYPLDHPLRKE 229  
QY 227 DEKHALLAGGHSRIIGPDGDLVAPLAENEGIIYANLDPCVRILAKMADPAGHYSR 296  
DB 230 HEGDKVILRGG--SCAVDPPLGTIVLEP-DFTKETIRYTFPLSDLAKKMDLVVGHYSR 286  
QY 297 PDITRLILDRSPKLPVVE 314  
DB 287 PDVQLKVNENSQSTVVK 304

RESULT 15  
T52266  
nitrilase-like protein [imported] - rice  
C:Species: Oryza sativa (rice)  
C>Date: 20-Oct-2000 #sequence\_revision 20-Oct-2000 #text\_change 31-Dec-2004  
C:Accession: T52266  
R:Chiba, R.; Dohmoto, M.; Yamaguchi, K.  
Submitted to the EMBL Data Library, May 1999  
A:Description: Oryza sativa, a gene for nitrilase-like protein.  
A:Reference number: Z26008

A:Accession: T52266  
 A:Status: preliminary; translated from GB/EMBL/DBJ  
 A:Molecule type: mRNA  
 A:Residues: 1-362 <CH1>  
 A:Cross-references: UNIPROT:Q9SXX6; UNIPARC:UPI000009C92A; EMBL:AB027054; PIDN:BAA77679.  
 A:Experimental source: cultivar Nipponbare  
 C:Genetics:  
 A:Gene: ONIT4  
 C:Superfamily: nitrilase (carbon-nitrogen hydrolase)

Query Match 25.6%; Score 451.5; DB 2; Length 362;  
 Best Local Similarity 33.5%; Pred. No. 3.4e-33;  
 Matches 109; Conservative 62; Mismatches 123; Indels 31; Gaps 7;

```

QY 5 IKVACVQAPFYMDLEATVVDKTIETMEBAARNNAFLAPPETWIPGP-----WFLW 56
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 31 VRATVVOASTVFDTPATLDKAERLIEBAAGYSQLVVFPFAFVGYPGSGTFPGANIS 90
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 57 LDSPA-WAMQFVROYHENSLELDGPQAKRISDAKRLGIMWTIGMSERVGGLYISQWFI 115
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 91 IGNPKDKGKEERKRYHAAIEVPGPEVTRILAMAGKYVFLVMGVIEREGTLLYCSVLFF 150
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 116 GDNGDTIGARRKLTPEVETLFGEGDSSLAVFETSVGRGLGCWEHLQPLTKYALYA 175
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 151 DPLGRYLGKHKRLMTALERITWGFQDSTIPYDTPIGKIGALICWENKMPDLRTALYG 210
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 176 QNEEIHCAAMPFSFLYPNAAKALGPDVNVVNASRIYAVGQCFTVLASCALVSQS----- 228
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 211 KGEIYYCA-----PTADSR---QVQASWTHALEGGCFVLSANQFCRKDYPPPE 259
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 229 -MIDMLCTDDEKHALLAGGHSRIIGPDGDLVAPLAENEHGLIYANLDPGVRIIAKMA 287
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 260 YVFTGLGEEPSDPDVVCPGG--SVLIISPSGEVLAGPNTG--EALITADLDLGEIVRAKFD 316
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 288 ADPAGHSRDPDITRLLIDRSPKLPV 312
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 317 FDVVGHYARPEVLSLVNDQPHLPV 341
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

Search completed: April 27, 2006, 01:01:07  
 Job time : 25.664 secs

**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: April 27, 2006, 00:50:35 ; Search time 151.477 Seconds  
(without alignments)  
1569.628 Million cell updates/sec

Title: US-09-751-299-4  
Perfect score: 1765  
Sequence: 1 MKEAIKVCVQAAPYIMDL.....DIRPYALGKASETGAQLBEI 337

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 2166443 seqs, 705528306 residues

Total number of hits satisfying chosen parameters: 2166443

Minimum DB seq length: 0  
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Uniprot 05.80:\*  
1: uniprot\_sprot:\*  
2: uniprot\_trembl:\*

Pred. No. is the number of results predicted by chance to have a  
score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1765	100.0	337	2	Q6RWE0_92222
2	1343	76.1	337	2	Q6RWE0_92222
3	1343	76.1	337	2	Q6RWE0_92222
4	1187	67.3	336	2	Q500U1_PSEBSY
5	1184	67.1	347	2	Q88B32_PSEBSM
6	1139	64.5	333	2	Q6RWE0_92222
7	1136	64.4	332	2	Q6RWE0_92222
8	1081	61.2	345	2	Q6RWE0_92222
9	1051	59.5	334	2	Q6RWE0_92222
10	1042	59.0	333	2	Q6RWE0_92222
11	1039	58.9	348	2	Q6RWE0_92222
12	1037	58.8	336	2	Q6RWE0_92222
13	1037	58.8	333	2	Q6RWE0_92222
14	1033	58.5	337	2	Q6RWE0_92222
15	1033	58.5	338	2	Q6RWE0_92222
16	1028.5	58.3	358	2	Q6RWE0_92222
17	1022.5	57.9	337	2	Q6RWE0_92222
18	1021.5	57.9	337	2	Q6RWE0_92222
19	1020.5	57.8	354	2	Q6RWE0_92222
20	1019.5	57.8	338	2	Q6RWE0_92222
21	1018	57.7	338	2	Q6RWE0_92222
22	1008	57.1	337	2	Q6RWE0_92222
23	1007	57.1	346	2	Q6RWE0_92222
24	999.5	56.6	338	2	Q6RWE0_92222
25	999	56.6	336	2	Q6RWE0_92222
26	997.5	56.5	338	2	Q6RWE0_92222
27	994.5	56.3	338	2	Q6RWE0_92222
28	993	56.3	338	2	Q6RWE0_92222
29	980.5	55.6	327	2	Q6RWE0_92222
30	976	55.3	352	2	Q6RWE0_92222
31	973.5	55.2	336	2	Q6RWE0_92222

32	971.5	55.0	341	2	Q6RWE0_92222	Q6RWE0_92222
33	937.5	53.1	335	2	Q6RWE0_92222	Q6RWE0_92222
34	936	53.0	331	2	Q6RWE0_92222	Q6RWE0_92222
35	929.5	52.7	337	2	Q6RWE0_92222	Q6RWE0_92222
36	924	52.4	350	2	Q6RWE0_92222	Q6RWE0_92222
37	919.5	52.1	334	2	Q6RWE0_92222	Q6RWE0_92222
38	918.5	52.0	334	2	Q6RWE0_92222	Q6RWE0_92222
39	914.5	51.8	337	2	Q6RWE0_92222	Q6RWE0_92222
40	906	51.3	345	2	Q6RWE0_92222	Q6RWE0_92222
41	898.5	50.9	349	2	Q6RWE0_92222	Q6RWE0_92222
42	881	49.9	346	2	Q6RWE0_92222	Q6RWE0_92222
43	836.5	47.4	365	1	NRL2_RHORN	Q03217 rhodococcus
44	819	46.4	356	1	NRL2_ALCPA	P20960 alcaligenes
45	800	45.3	383	1	NRL1_RHORN	Q02068 rhodococcus

## ALIGNMENTS

RESULT 1					
Q6RWE0_92222 PRELIMINARY; PRT: 337 AA.					
ID	Q6RWE0_92222	Q6RWE0_92222	Q6RWE0_92222	Q6RWE0_92222	Q6RWE0_92222
AC	Q6RWE0_92222	Q6RWE0_92222	Q6RWE0_92222	Q6RWE0_92222	Q6RWE0_92222
DT	05-JUL-2004 (TrEMBLrel. 27, Created)				
DT	05-JUL-2004 (TrEMBLrel. 27, Last sequence update)				
DT	05-JUL-2004 (TrEMBLrel. 27, Last annotation update)				
DE	Nitriase (EC 3.5.5.7).				
GN	ORFNames=BD5220;				
OS	uncultured organism.				
OC	unclassified; environmental samples.				
OX	NCBI_TaxID=155900;				
RN	[1]				
RP	NUCLEOTIDE SEQUENCE.				
RX	PubMed=15066841; DOI=10.1128/JEM.70.4.2429-2436.2004;				
RA	Robertson D.E., Chaplin J.A., Desantis G., Podar M., Madden M.,				
RA	Chi E., Richardson T., Milan A., Miller M., Weiner D.P., Wong K.,				
RA	McQuaid J., Farwell B., Preston L.A., Tan X., Sneed M.A., Keller M.,				
RA	Mathur E., Kretz P.L., Burk M.J., Short J.M.;				
RT	"Exploring nitriase sequence space for enantioselective catalysts.";				
RL	Appl. Environ. Microbiol. 70:2429-2436(2004).				
DR	EMBL; AY487562; AAR97509.1; -; Genomic DNA.				
DR	GO; GO:0018762; F:aliphatic nitriase activity; IEA.				
DR	GO; GO:0016810; F:hydrolase activity; acting on carbon-nitrog. . .; IEA.				
DR	GO; GO:0006807; P:nitrogen compound metabolism; IEA.				
DR	InterPro; IPR00132; N/CN hydrolase.				
DR	InterPro; IPR003010; N/CN hydrolase.				
DR	Pfam; PF00795; CN hydrolase; 1.				
DR	PROSITE; PS00263; CN HYDROLASE; 1.				
KW	PROSITE; PS00921; NITRIL_CHT_2; 1.				
KW	Hydrolase. 337 AA; 36783 MW; 25D2DE1B820983B2 CRC64;				
SQ SEQUENCE					
Query Match					
Best Local Similarity 100.0%; Pred. No. 4.9e-145;					
Matches 337; Conservative 0; Mismatches 0; Indels 0; Gaps 0;					
QY	1	MKEAIKVCVQAAPYIMDL	1	Q6RWE0_92222	Q6RWE0_92222
DB	1	MKEAIKVCVQAAPYIMDL	1	Q6RWE0_92222	Q6RWE0_92222
QY	61	AMAMQFQVYHENSLELGPQAKRISDAKXGLGIVNTLGMSEBVGCTLYISQWFTGDND	61	Q6RWE0_92222	Q6RWE0_92222
DB	61	AMAMQFQVYHENSLELGPQAKRISDAKXGLGIVNTLGMSEBVGCTLYISQWFTGDND	61	Q6RWE0_92222	Q6RWE0_92222
QY	121	TIGARRKLKPTFVERTLFGEDGSSLAFTSVGRIGLACMEHLQPLTKYALYAQNEI	121	Q6RWE0_92222	Q6RWE0_92222
DB	121	TIGARRKLKPTFVERTLFGEDGSSLAFTSVGRIGLACMEHLQPLTKYALYAQNEI	121	Q6RWE0_92222	Q6RWE0_92222
QY	181	HCAAMPSSSLYPNAKALGPVNTAASRTIYAVEGQCPVLASCAVVSQSMIMLCTDDEKH	181	Q6RWE0_92222	Q6RWE0_92222
DB	181	HCAAMPSSSLYPNAKALGPVNTAASRTIYAVEGQCPVLASCAVVSQSMIMLCTDDEKH	181	Q6RWE0_92222	Q6RWE0_92222
QY	240	ALLAGGSHIIGPDGDDVAPLAENEGILVNLNDGVRILAKMADPAGHTSRPDT	240	Q6RWE0_92222	Q6RWE0_92222
DB	240	ALLAGGSHIIGPDGDDVAPLAENEGILVNLNDGVRILAKMADPAGHTSRPDT	240	Q6RWE0_92222	Q6RWE0_92222

QY	DB	301	RLILDRSPKLVPEVEIGDLRPYALGKASETGAQLEEI	337
DB	241	ALLTAGGSHSIIIGDGDVLPAFLAENEGLIYANLDPGVAILAKMAADPAHYSRPDIT	300	
QY	301	RLILDRSPKLVPEVEIGDLRPYALGKASETGAQLEEI	337	
DB	301	RLILDRSPKLVPEVEIGDLRPYALGKASETGAQLEEI	337	

RESULT 2

Q6RMQ3 92ZZZ PRELIMINARY; PRT; 337 AA.

AC Q6RMQ3-

DT 05-JUL-2004 (TREMBlrel. 27, Created)

DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)

DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)

DE Nitrilase (EC 3.5.5.7).

GN ORFNames=BD5275;

OS unclutured organism.

OC unclassified: environmental samples.

OX NCBI\_Taxid=155900;

RN (1)

RP NUCLEOTIDE SEQUENCE.

RA PubMed=15066841; DOI=10.1128/AEM.70.4.2429-2436.2004;

RA Robertson D.E., Chaplin J.A., Desantis G., Podar M., Madden M.,

RA Chi E., Richardson T., Milan A., Miller M., Weiner D.P., Wong K.,

RA McDonald J., Farwell B., Preston L.A., Tan X., Sneed M.A., Keller M.,

RA Mathur E., Kretz P.L., Burk M.J., Short J.M.;

RT "Exploring nitrilase sequence space for enantioselective catalysts";

RT Appl. Environ. Microbiol. 70:2429-2436(2004).

EMBL: AY487449; AAR97396.1; -; Genomic\_DNA.

DR GO; GO:0016810; F:aliphatic nitrilase activity; IEA.

DR GO; GO:0006807; P:nitrogenase activity, acting on carbon-nitrog. .; IEA.

DR InterPro: IPR000132; N/CN hydrolase.

DR InterPro: IPR003010; Ntlase/CNhydase.

DR Pfam: PF00795; CN\_hydrolase; 1.

DR PROSITE: PS50263; CN\_HYDROLASE; 1.

DR PROSITE: PS00920; NITRIL\_CHT\_1; 1.

DR PROSITE: PS00921; NITRIL\_CHT\_2; 1.

KW Hydrolase.

SO SEQUENCE 337 AA; 36697 MW; 940993D5AB69582E CRC64;

QY	DB	301	RLILDRSPKLVPEVEIGDLRPYALGKASETGAQLEEI	337
QY <td>301 <td>RLILDRSPKLVPEVEIGDLRPYALGKASETGAQLEEI <td>337</td> <td></td> </td></td>	301 <td>RLILDRSPKLVPEVEIGDLRPYALGKASETGAQLEEI <td>337</td> <td></td> </td>	RLILDRSPKLVPEVEIGDLRPYALGKASETGAQLEEI <td>337</td> <td></td>	337	
DB	301 <td>RLILDRSPKLVPEVEIGDLRPYALGKASETGAQLEEI <td>337</td> <td></td> </td>	RLILDRSPKLVPEVEIGDLRPYALGKASETGAQLEEI <td>337</td> <td></td>	337	

RESULT 3

Q6RMQ7 92ZZZ

ID	Query Match	Best Local Similarity	Matches 253;	Conservative 29;	Mismatches 478;	Indels 6;	Gaps 1
1D	Q6KW07_92ZZZ PRELIMINARY;	PRT; 337 AA.					
AC	Q6KW07_						
DT	05-JUL-2004 (TREMBlrel. 27, Created)						
DT	05-JUL-2004 (TREMBlrel. 27, Last sequence update)						
DT	05-JUL-2004 (TREMBlrel. 27, Last annotation update)						
DE	Nitrilase (EC 3.5.5.7).						
CN	ORFNames=BD5274;						
OS	uncultured organism.						
OC	Unclassified environmental samples.						
OX	NCBI_TaxID=155900;						
RP	NUCLEOTIDE SEQUENCE.						
RA	Published:1506641; DOI=10.1128/AEM.70.4.2429-2436.2004;						
RA	Robertson D.E., Chaplin J.A., Desantis G., Podar M., Madden M.,						
RA	Chi E., Richardson T., Milan A., Miller M., Weiner D.P., Wong K.,						
RA	McQuaid J., Fairwell B., Preston L.A., Tan X., Snead M.A., Keller M.,						
RA	Matthur E., Kretz P.L., Burk M.J., Short J.M.,						
RT	"Exploring nitrilase sequence space for enantioselective catalysis."						
RL	Appl. Environ. Microbiol. 70:2429-2436(2004).						
DR	EMBL: AY487445; AAR97392.1; -; Genomic DNA.						
DR	GO: 0016810; F:hydrolyase activity, acting on carbon-nitrog. .; IEA.						
DR	GO: 00106807; F:nitrogen compound metabolism; IEA.						
DR	InterPro: IPR00132; N/CN hydrolase.						
DR	InterPro: IPR00795; CN hydrolase; 1.						
DR	Pfam: PF00263; CN HYDROLASE; 1.						
DR	PROSITE: PS00920; NITRIL CHT 1; 1.						
DR	PROSITE: PS00921; NITRIL_CHT_2; 1.						
KW	Hydrolase. 337 AA, 36769 MW; 9409905AB69582E CRC64;						
SQ	SEQUENCE.						
QY	1 MEAIKVVACVQAAPPIYMDLEAVDVTIELMEERANRNRLIAFPETWIPGYWPFILDSP 60						
DB	1 MEAIKVVACVQAAPFPLDIDAVDKVALLIEAARNGARLIFPETWIPGYWPFILDSP 60						
QY	61 AWAAMQFVROYHNSLELDGPQAKRISDAKRLGIMWTIGMSERVGGLIYSQWFIQDNGD 120						
DB	61 ANGMQFVRRIYHNSLVLDSPQAKRISDAQAGRIYVALGYSRVSGLTYMGOMLIDDKGE 120						
QY	121 TTGARRKLPFTVERTLFEFGGSSLAIVETSVGNLGLICMEHHOPLTKVALLYQNEST 180						
DB	121 TAGIRRKLPFTVERTLFEFGGSSLSITDPTLGVGLGCMEHHOPLSKVALLYQNEST 180						
QY	181 HCAWPSFSLPYPAALGPDVNVVAASRIYAVEGQCFVLASCALYSQSMIDMLCTDDEKH 240						
DB	181 HPAWPSFSLPYPAALGPDVNVVAASRIYAVEGQCFVLASCALYSPEMIEMLCTDESKH 240						
QY	241 ALLIAGGHSRIITGPDGDLVAPLAENEGILYANIDPGVRIIAKKAADPAGHSRDLT 300						
DB	241 SLIAGGGSRIITGPDGSLAPLIGNEGIIYATIDPARIYAKTAADPAGHSRDLT 300						
QY	301 RLIIIDSPKLPVEIEGDLRPVALGKASENGAOLE 335						
DB	301 RLIIINSAQPVVEVGREI-----PASAQGEFVE 329						
RESULT 4							
OS0001_PSESY							
ID	Q500U1_PSESY PRELIMINARY;	PRT; 336 AA.					
AC	Q500U1;						
DT	13-SEP-2005 (TREMBlrel. 31, Created)						
DT	13-SEP-2005 (TREMBlrel. 31, Last sequence update)						
DT	13-SEP-2005 (TREMBlrel. 31, Last annotation update)						
DE	Aliphatic nitrilase (EC 3.5.5.7).						
CN	ORFNames=PSyr_0007;						
OS	Pseudomonas syringae pv. syringae B728a.						
OC	Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;						
OC	Pseudomonadaceae; Pseudomonas.						

RX NCBI\_TaxID=205918;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=B728a;  
 RG DOE Joint Genome Institute;  
 RA Chain P., Larimer F., Dibartolo G., Copeland A., Lykidis A., Truong S.,  
 RA Nolan M., Goldsman E., Thiel J., Malfatti S., Lapidus A., Detter J.C.,  
 RA Land M., Richardson P.M., Kyrpides N.C., Ivanova N.,  
 RT "Comparison of two complete genome sequences of *Pseudomonas syringae*  
 RT pv. *syringae* B728a and pv. *tomato* DC3000.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 0:0-0(2005).  
 RN [2]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=B728a;  
 RL Loper J.;  
 RN Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.  
 RN [3]  
 RP NUCLEOTIDE SEQUENCE.  
 RC STRAIN=B728a;  
 RA Fell H., Fell W.S., Lindow S.E.;  
 RL Submitted (MAY-2005) to the EMBL/GenBank/DBJ databases.  
 DR EMBL: CP000075; AA035081.1; -; Genomic\_DNA.  
 DR GO: GO:0018762; F:aliphatic nitrilase activity; IEA.  
 DR GO: GO:0016787; F:hydrolase activity; IEA.  
 KW Hydrolase.  
 SQ SEQUENCE 336 AA; 36552 MW; EC3DF2A762B37729 CRC64;

Query Match 67.3%; Score 1187; DB 2; Length 336;  
 Best Local Similarity 69.2%; Pred. No. 1,1e-94;  
 Matches 218; Conservative 37; Mismatches 60; Indels 0; Gaps 0;

QY 1 MKEAIKACVQAAYIVMDLEATVDTKTELMEEAARNRRLIAPETWIPGYPWFLMDSP 60  
 DB 1 MKEPLKACVQAAPVFLDLDATVDTKTLTLMQAAAGAGLAFETWIPGYPWFLMDAP 60  
 QY 61 AMAMQFROYHENSLELDGPOAKRISDAKRLGIMVTLGMSERVGTLTYSQWFIGNGD 120  
 DB 61 AMNPLVORHYQSLVLDVQARRISDAARHGLYVVLGYSERKASLYIGQWIIDHGE 120  
 QY 121 TIGARRKLKPTFVTRTLFGSDSSSLAVFETSVGRGLGCWEHLQPLTKALYAQNEEI 180  
 DB 121 TVGTRRLKATHTVTRTLFGSDGASLRFETPVGVLGALCWEHLQPLSKYAMTAQNEOI 180  
 QY 181 HCAWPSFSLYPNAKALGPVNVAAASRIYAVEGQCFVLASCVLSQSMIDMLCTDDEKH 240  
 DB 181 HVAWPSFSLYRNATSLGPEVNTAASRYVAAGQCFVLAPCAIVSPBTEMLCDSDAKR 240  
 QY 241 ALLAGGSHRIIGPDGDLVAPLAENEGILYANLDPGVRIILAKMAADPAHYSRDIT 300  
 DB 241 SLQAGGSHARIRPGDSDLATPLGHEHBEGLVATLDPALILAKVAADPAHYSRDIT 300  
 QY 301 RLIDRSFKLPVIEI 315  
 DB 301 RLMEFNPPTPCVVDL 315

RESULT 5  
 Q88B32\_PSESM PRELIMINARY; PRT; 347 AA.  
 AC Q88B32;  
 DT 01-JUN-2003 (TREMBlrel. 24, Created)  
 DT 01-JUN-2003 (TREMBlrel. 24, Last sequence update)  
 DT 01-OCT-2003 (TREMBlrel. 25, Last annotation update)  
 DE Nitrilase, putative.  
 GN OrderedLocustNames=PSPT00189;  
 OS *Pseudomonas syringae* (pv. *tomato*).  
 OC Bacteria; Proteobacteria; Gammaproteobacteria; Pseudomonadales;  
 OC Pseudomonadaceae; *Pseudomonas*.  
 OX NCBI\_TaxID=323;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE (LARGE SCALE GENOMIC DNA).  
 RC STRAIN=DC3000;  
 RX MEDLINE=22834015; PubMed=12928499; DOI=10.1073/pnas.1731982100;

RA Buell C.R., Joardar V., Lindeberg M., Selengut J., Paulsen I.T.,  
 RA Gwin M.L., Dodson R.J., DeBoy R.T., Durkin A.S., Kolonay J.F.,  
 RA Madupu R., Daugherty S.C., Brinkac L.M., Beaman M.J., Haft D.H.,  
 RA Nelson W.C., Davidson T.M., Zafar N., Zhou L., Liu J., Yuan Q.,  
 RA Khouri H.M., Fedorova N.B., Tran B., Russell D., Berry K.J.,  
 RA Utechtack T.R., Van Aken S.E., Feldblyum T.V., D'Ascenzo M.,  
 RA Deng W.-L., Ramos A.R., Alfano J.R., Carlinhour S., Chatterjee A.K.,  
 RA Delaney T.P., Lazarewitz S.G., Martin G.B., Schneider D.J., Tang X.,  
 RA Bender C.L., White O., Fraser C.M., Collier A.;  
 RT "The complete genome sequence of the Arabidopsis and tomato pathogen  
 RT *Pseudomonas syringae* pv. *tomato* DC3000.";  
 RL Proc. Natl. Acad. Sci. U.S.A. 100:10181-10186(2003).  
 DR EMBL: AB016853; AA053742.1; -; Genomic\_DNA.  
 DR TIGR: PSP00189; -;  
 DR GO: GO:0016810; F:hydrolase activity; acting on carbon-nitrog. . .; IEA.  
 DR GO: GO:0006807; P:nitrogen compound metabolism; IEA.  
 DR InterPro: IPR000132; N/CN\_hydrolase.  
 DR InterPro: IPR003010; N/ase/CNhydrolase.  
 DR Pfam: PF00795; CN\_hydrolase; 1.  
 DR PROSITE: PS50263; CN\_HYDROLASE; 1.  
 DR PROSITE: PS00920; NITRIL\_CHT\_1; 1.  
 DR PROSITE: PS00921; NITRIL\_CHT\_2; 1.  
 KW Complete proteome.  
 SQ SEQUENCE 347 AA; 37564 MW; B671F783D7411308 CRC64;

Query Match 67.1%; Score 1184; DB 2; Length 347;  
 Best Local Similarity 68.8%; Pred. No. 2e-94;  
 Matches 218; Conservative 37; Mismatches 62; Indels 0; Gaps 0;

QY 1 MKEAIKACVQAAYIVMDLEATVDTKTELMEEAARNRRLIAPETWIPGYPWFLMDSP 60  
 DB 12 MKEPLKACVQAAPVFLDLDATVDTKTLTLMQAAAGAGLAFETWIPGYPWFLMDAP 71  
 QY 61 AMAMQFROYHENSLELDGPOAKRISDAKRLGIMVTLGMSERVGTLTYSQWFIGNGD 120  
 DB 72 AMNPLVORHYQSLVLDVQARRISDAARHGLYVVLGYSERKASLYIGQWIIDHGE 131  
 QY 121 TIGARRKLKPTFVTRTLFGSDSSSLAVFETSVGRGLGCWEHLQPLTKALYAQNEEI 180  
 DB 132 TVGTRRLKATHTVTRTLFGSDGASLRFETPVGVLGALCWEHLQPLSKYAMTAQNEOI 191  
 QY 181 HCAWPSFSLYPNAKALGPVNVAAASRIYAVEGQCFVLASCVLSQSMIDMLCTDDEKH 240  
 DB 192 HVAWPSFSLYQVATRALGPEVNTAASRYVAAGQCFVLAPCAIVSPBTEMLCDSDAKH 251  
 QY 241 ALLAGGSHRIIGPDGDLVAPLAENEGILYANLDPGVRIILAKMAADPAHYSRDIT 300  
 DB 252 SLQAGGSHARIRPGDSDLATPLGHEHBEGLVATLDPALILAKVAADPAHYSRDIT 311  
 QY 301 RLIDRSFKLPVIEI 317  
 DB 312 RLMEFNPPTPCVVDL 328

RESULT 6  
 Q6RWLO\_92ZZZ PRELIMINARY; PRT; 333 AA.  
 AC Q6RWLO;  
 DT 05-JUL-2004 (TREMBlrel. 27, Created)  
 DT 05-JUL-2004 (TREMBlrel. 27, Last sequence update)  
 DT 05-JUL-2004 (TREMBlrel. 27, Last annotation update)  
 DE Nitrilase (EC 3.5.5.7).  
 GN ORFNames=BD7615;  
 OS uncultured organism.  
 OC unclassified; environmental samples.  
 OX NCBI\_TaxID=155900;  
 RN [1]  
 RP NUCLEOTIDE SEQUENCE.  
 RX PubMed=1506841; DOI=10.1126/ASM.70.4.2429-2436.2004;  
 RA Robertson D.E., Chaplin J.A., Desantis G., Podar M., Madden M.,  
 RA Chi E., Richardson T., Milan A., Miller M., Weiner D.P., Wong K.,  
 RA McQuaid J., Farwell B., Preston L.A., Tan X., Sneed M.A., Keller M.,  
 RA Methur E., Kretz P.L., Burk M.J., Short J.M.;



```
Db 9 RVAAQAAPERLNTAEVDTKITALIEBARAGASLIAPETWIPGYPMFAMLGAPIMGAK 68
Qy 66 FVRQYHNSLELDGQAKRISDAARLIGIMWTLMGSEVGGTLYISQWFIGDNGCTIGAR 125
Db 69 FIAQAHDSMSWIDGQAFERIAQAARCNITVVLGFSEKDGSLYIAQAILSPGKTIATYR 128
Qy 126 RKLKPTFERTLFGEGDSSLAVFETSVGRGLGCMWEHLOPLTYALYAONEEIHCAW 185
Db 129 RKLKPTFAERLIFGEGDSSDLAVHDTKIGRVGALCCMEHLOPLSKYAMTAQNEOVHIAW 188
Qy 186 PSFSLYPNAKALGPDVNVAAASRIYAVEGQCFVLASCALVSQSMIDMLCTDDEKHALLA 245
Db 189 PSFSLYVDAAYALGPEVNNAASRLIYAVEGQCFVVAPCAIVSQKIMIDMLCEPPEQALLKP 248
Qy 246 GGGHSRITIGPDGDDVAPLAENEGILYANLDPGVRIILAKMAADPAGHSRPDITRL 305
Db 249 GGGHAQIYGPDRSLADLPDABGLIYADIDLAIITLAKAAMDPAHGHSRPDVTQLLLD 308
Qy 306 RSPKLPVV 313
Db 309 RNPKRNV 316

RESULT 9
Q6RWF9_9222Z PRELIMINARY; PRT; 353 AA.
ID Q6RWF9_9222Z PRELIMINARY; PRT; 353 AA.
AC Q6RWF9;
RX PubMed15066841; DOI=10.1128/AEM.70.4.2429-2436.2004;
RA Robertson D.E., Chaplin J.A., Desantis G., Podar M., Madden M.,
RA Chi E., Richardson T., Milon A., Miller M., Weiner D.P., Wong K.,
RA McQuaid J., Farrell B., Preston L.A., Tan X., Sneed M.A., Keller M.,
RA Mathur E., Kretz P.L., Burk M.J., Short J.M.;
RT "Exploring nitrilase sequence space for enantioselective catalysis.";
RL Appl. Environ. Microbiol. 70:2429-2436(2004).
DR EMBL; AY487559; AAR97506.1; -; Genomic DNA.
DR GO; GO:0018762; F:aliphatic nitrilase activity; IEA.
DR GO; GO:0006807; P:nitrogen compound metabolism; IEA.
DR InterPro; IPR000132; N/CN_hydrolase.
DR InterPro; IPR003010; NClase/CNhydase.
DR Pfam; PF00795; CN_hydrolase; 1.
DR PROSITE; PS50263; CN_HYDROLASE; 1.
DR PROSITE; PS00920; NITRIL_CHT_1; 1.
DR PROSITE; PS00921; NITRIL_CHT_2; 1.
KW Hydrolase.
SQ SEQUENCE 353 AA; 38029 MW; 997D13779FC31326 CRC64;

Query Match 59.0%; Score 1042; DB 2; Length 353;
Best Local Similarity 63.6%; Pred. No. 4.7e-82;
Matches 194; Conservative 44; Mismatches 65; Indels 2; Gaps 1;
```

```
Qy 181 HCAAPSFSLYPNAKALGPDVNVAAASRIYAVEGQCFVLASCALVSQSMIDMLCTDDEKH 240
Db 187 HVAAMPSEFSPKPEPAHALGMEVNNAASKIYAVESGCFVLGACAYISQWAMDEMCDTEDKR 246
Qy 241 ALLLAGGHSRIIGPDGDDVAPLAENEGILYANLDPGVRIILAKMAADPAGHSRPDIT 300
Db 247 ALVHAGGGHNAVIFPDGRSLADKIPETQEGILYADIDLGAIVAKMAADPAGHSRPDVT 306
Qy 301 RLIDRSP-----KLPPVEIEGDLRP 321
Db 307 RLFPNNKPARREVEFSLPVDATETPPQ 334

RESULT 10
Q6RWF9_9222Z PRELIMINARY; PRT; 353 AA.
ID Q6RWF9_9222Z PRELIMINARY; PRT; 353 AA.
AC Q6RWF9;
RX PubMed15066841; DOI=10.1128/AEM.70.4.2429-2436.2004;
RA Robertson D.E., Chaplin J.A., Desantis G., Podar M., Madden M.,
RA Chi E., Richardson T., Milon A., Miller M., Weiner D.P., Wong K.,
RA McQuaid J., Farrell B., Preston L.A., Tan X., Sneed M.A., Keller M.,
RA Mathur E., Kretz P.L., Burk M.J., Short J.M.;
RT "Exploring nitrilase sequence space for enantioselective catalysis.";
RL Appl. Environ. Microbiol. 70:2429-2436(2004).
DR EMBL; AY487543; AAR97490.1; -; Genomic DNA.
DR GO; GO:0018762; F:aliphatic nitrilase activity; IEA.
DR GO; GO:0006807; P:nitrogen compound metabolism; IEA.
DR InterPro; IPR000132; N/CN_hydrolase.
DR InterPro; IPR003010; NClase/CNhydase.
DR Pfam; PF00795; CN_hydrolase; 1.
DR PROSITE; PS50263; CN_HYDROLASE; 1.
DR PROSITE; PS00920; NITRIL_CHT_1; 1.
DR PROSITE; PS00921; NITRIL_CHT_2; 1.
KW Hydrolase.
SQ SEQUENCE 353 AA; 38029 MW; 997D13779FC31326 CRC64;

Query Match 59.0%; Score 1042; DB 2; Length 353;
Best Local Similarity 63.6%; Pred. No. 4.7e-82;
Matches 194; Conservative 44; Mismatches 65; Indels 2; Gaps 1;
```

[illegible]

Query	Best Local Match	Similarity	Score	DB 2	Length	336
Matches	197	Conservative	46	Mismatches	73	Indels
05-JUL-2004 (TREMBlrel. 27, Created)	05-JUL-2004 (TREMBlrel. 27, Last sequence update)	05-JUL-2004 (TREMBlrel. 27, Last sequence update)	05-JUL-2004 (TREMBlrel. 27, Last sequence update)	05-JUL-2004 (TREMBlrel. 27, Last sequence update)	05-JUL-2004 (TREMBlrel. 27, Last sequence update)	05-JUL-2004 (TREMBlrel. 27, Last sequence update)
DE Nitrilase (EC 3.5.5.7).	DE Nitrilase (EC 3.5.5.7).	DE Nitrilase (EC 3.5.5.7).	DE Nitrilase (EC 3.5.5.7).	DE Nitrilase (EC 3.5.5.7).	DE Nitrilase (EC 3.5.5.7).	DE Nitrilase (EC 3.5.5.7).
ORFNames=BD5295;	ORFNames=BD5295;	ORFNames=BD5295;	ORFNames=BD5295;	ORFNames=BD5295;	ORFNames=BD5295;	ORFNames=BD5295;
OS uncultured organism.	OS uncultured organism.	OS uncultured organism.	OS uncultured organism.	OS uncultured organism.	OS uncultured organism.	OS uncultured organism.
OC unclassified; environmental samples.	OC unclassified; environmental samples.	OC unclassified; environmental samples.	OC unclassified; environmental samples.	OC unclassified; environmental samples.	OC unclassified; environmental samples.	OC unclassified; environmental samples.
OX NCBI_Taxid=155900;	OX NCBI_Taxid=155900;	OX NCBI_Taxid=155900;	OX NCBI_Taxid=155900;	OX NCBI_Taxid=155900;	OX NCBI_Taxid=155900;	OX NCBI_Taxid=155900;
1]	1]	1]	1]	1]	1]	1]
NCLECTED SEQUENCE.	NCLECTED SEQUENCE.	NCLECTED SEQUENCE.	NCLECTED SEQUENCE.	NCLECTED SEQUENCE.	NCLECTED SEQUENCE.	NCLECTED SEQUENCE.
RP PubMed=15066841, DOI=10.1128/AEM.70.4.2429-2436.2004;	RP PubMed=15066841, DOI=10.1128/AEM.70.4.2429-2436.2004;	RP PubMed=15066841, DOI=10.1128/AEM.70.4.2429-2436.2004;	RP PubMed=15066841, DOI=10.1128/AEM.70.4.2429-2436.2004;	RP PubMed=15066841, DOI=10.1128/AEM.70.4.2429-2436.2004;	RP PubMed=15066841, DOI=10.1128/AEM.70.4.2429-2436.2004;	RP PubMed=15066841, DOI=10.1128/AEM.70.4.2429-2436.2004;
RA Robertson D.S., Chaplin J.A., Desantis G., Podar M., Madden M.,	RA Robertson D.S., Chaplin J.A., Desantis G., Podar M., Madden M.,	RA Robertson D.S., Chaplin J.A., Desantis G., Podar M., Madden M.,	RA Robertson D.S., Chaplin J.A., Desantis G., Podar M., Madden M.,	RA Robertson D.S., Chaplin J.A., Desantis G., Podar M., Madden M.,	RA Robertson D.S., Chaplin J.A., Desantis G., Podar M., Madden M.,	RA Robertson D.S., Chaplin J.A., Desantis G., Podar M., Madden M.,
RA Cui E., Richardson T., Milan A., Miller M., Weiner D.P., Wong K.,	RA Cui E., Richardson T., Milan A., Miller M., Weiner D.P., Wong K.,	RA Cui E., Richardson T., Milan A., Miller M., Weiner D.P., Wong K.,	RA Cui E., Richardson T., Milan A., Miller M., Weiner D.P., Wong K.,	RA Cui E., Richardson T., Milan A., Miller M., Weiner D.P., Wong K.,	RA Cui E., Richardson T., Milan A., Miller M., Weiner D.P., Wong K.,	RA Cui E., Richardson T., Milan A., Miller M., Weiner D.P., Wong K.,
RA McQuaid J., Farrell B., Preston L.A., Tan X., Sneed M.A., Keller M.,	RA McQuaid J., Farrell B., Preston L.A., Tan X., Sneed M.A., Keller M.,	RA McQuaid J., Farrell B., Preston L.A., Tan X., Sneed M.A., Keller M.,	RA McQuaid J., Farrell B., Preston L.A., Tan X., Sneed M.A., Keller M.,	RA McQuaid J., Farrell B., Preston L.A., Tan X., Sneed M.A., Keller M.,	RA McQuaid J., Farrell B., Preston L.A., Tan X., Sneed M.A., Keller M.,	RA McQuaid J., Farrell B., Preston L.A., Tan X., Sneed M.A., Keller M.,
RA Mathur E., Kreitz P.L., Burk M.J., Short J.M.;	RA Mathur E., Kreitz P.L., Burk M.J., Short J.M.;	RA Mathur E., Kreitz P.L., Burk M.J., Short J.M.;	RA Mathur E., Kreitz P.L., Burk M.J., Short J.M.;	RA Mathur E., Kreitz P.L., Burk M.J., Short J.M.;	RA Mathur E., Kreitz P.L., Burk M.J., Short J.M.;	RA Mathur E., Kreitz P.L., Burk M.J., Short J.M.;
RA "Exploring nitrilase sequence space for enantioselective catalysis.";	RA "Exploring nitrilase sequence space for enantioselective catalysis.";	RA "Exploring nitrilase sequence space for enantioselective catalysis.";	RA "Exploring nitrilase sequence space for enantioselective catalysis.";	RA "Exploring nitrilase sequence space for enantioselective catalysis.";	RA "Exploring nitrilase sequence space for enantioselective catalysis.";	RA "Exploring nitrilase sequence space for enantioselective catalysis.";
RL Appl. Environ. Microbiol. 70:2429-2436(2004).	RL Appl. Environ. Microbiol. 70:2429-2436(2004).	RL Appl. Environ. Microbiol. 70:2429-2436(2004).	RL Appl. Environ. Microbiol. 70:2429-2436(2004).	RL Appl. Environ. Microbiol. 70:2429-2436(2004).	RL Appl. Environ. Microbiol. 70:2429-2436(2004).	RL Appl. Environ. Microbiol. 70:2429-2436(2004).
DR EMBL, AY487546; AAR97493.1; -; Genomic DNA.	DR EMBL, AY487546; AAR97493.1; -; Genomic DNA.	DR EMBL, AY487546; AAR97493.1; -; Genomic DNA.	DR EMBL, AY487546; AAR97493.1; -; Genomic DNA.	DR EMBL, AY487546; AAR97493.1; -; Genomic DNA.	DR EMBL, AY487546; AAR97493.1; -; Genomic DNA.	DR EMBL, AY487546; AAR97493.1; -; Genomic DNA.
DR GO; GO:0018762; F:aliphatic nitrilase activity; IEA.	DR GO; GO:0018762; F:aliphatic nitrilase activity; IEA.	DR GO; GO:0018762; F:aliphatic nitrilase activity; IEA.	DR GO; GO:0018762; F:aliphatic nitrilase activity; IEA.	DR GO; GO:0018762; F:aliphatic nitrilase activity; IEA.	DR GO; GO:0018762; F:aliphatic nitrilase activity; IEA.	DR GO; GO:0018762; F:aliphatic nitrilase activity; IEA.
DR GO; GO:0016810; F:hydrolase activity, acting on carbon-nitrog. . . ; IEA.	DR GO; GO:0016810; F:hydrolase activity, acting on carbon-nitrog. . . ; IEA.	DR GO; GO:0016810; F:hydrolase activity, acting on carbon-nitrog. . . ; IEA.	DR GO; GO:0016810; F:hydrolase activity, acting on carbon-nitrog. . . ; IEA.	DR GO; GO:0016810; F:hydrolase activity, acting on carbon-nitrog. . . ; IEA.	DR GO; GO:0016810; F:hydrolase activity, acting on carbon-nitrog. . . ; IEA.	DR GO; GO:0016810; F:hydrolase activity, acting on carbon-nitrog. . . ; IEA.
DR GO; GO:0006807; P:nitrogen compound metabolism; IEA.	DR GO; GO:0006807; P:nitrogen compound metabolism; IEA.	DR GO; GO:0006807; P:nitrogen compound metabolism; IEA.	DR GO; GO:0006807; P:nitrogen compound metabolism; IEA.	DR GO; GO:0006807; P:nitrogen compound metabolism; IEA.	DR GO; GO:0006807; P:nitrogen compound metabolism; IEA.	DR GO; GO:0006807; P:nitrogen compound metabolism; IEA.
DR InterPro; IPR001032; N:CN hydrolase.	DR InterPro; IPR001032; N:CN hydrolase.	DR InterPro; IPR001032; N:CN hydrolase.	DR InterPro; IPR001032; N:CN hydrolase.	DR InterPro; IPR001032; N:CN hydrolase.	DR InterPro; IPR001032; N:CN hydrolase.	DR InterPro; IPR001032; N:CN hydrolase.
DR InterPro; IPR001032; N:CN hydrolase.	DR InterPro; IPR001032; N:CN hydrolase.	DR InterPro; IPR001032; N:CN hydrolase.	DR InterPro; IPR001032; N:CN hydrolase.	DR InterPro; IPR001032; N:CN hydrolase.	DR InterPro; IPR001032; N:CN hydrolase.	DR InterPro; IPR001032; N:CN hydrolase.
DR Pfam; PF00795; CN hydrolase; 1.	DR Pfam; PF00795; CN hydrolase; 1.	DR Pfam; PF00795; CN hydrolase; 1.	DR Pfam; PF00795; CN hydrolase; 1.	DR Pfam; PF00795; CN hydrolase; 1.</		

RA Robertson D.E., Chaplin J.A., Desantis G., Podar M., Madden M.,  
RA Chi E., Richardson T., Milan A., Miller M., Weiner D.P., Wong K.,  
RA McQuaid J., Farwell B., Preston L.A., Tan X., Sneed M.A., Keller M.,  
RA Mathur E., Kretz P.L., Burk M.J., Short J.M.,  
RT "Exploring nitrilase sequence space for enantioselective catalysis.",  
RL Appl. Environ. Microbiol. 70:2429-2436(2004).  
DR EMBL; AY487552; AAR97499.1; -; Genomic DNA.  
DR GO; GO:0018762; F:aliphatic nitrilase activity; IEA.  
DR GO; GO:0016810; F:hydrolyase activity, acting on carbon-nitrog. .; IEA.  
DR GO; GO:0006807; P:nitrogen compound metabolism; IEA.  
DR InterPro; IPR000132; N/CN hydrolase.  
DR InterPro; IPR003010; Ntlase/CNhydase.  
DR Pfam; PF00795; CN hydrolase; 1.  
DR PROSITE; PS50263; CN HYDROLASE; 1.  
DR PROSITE; PS00920; NITRIL\_CHT\_1; 1.  
DR PROSITE; PS00921; NITRIL\_CHT\_2; 1.  
DR Hydrolyase.  
SQ SEQUENCE 353 AA; 38321 MW; 2AA05257B3A7BC6C CRC64;  
  
Query Match 58.8%; Score 1037; DB 2; Length 353;  
Best Local Similarity 59.5%; Pred. No. 1.3e-81;  
Matches 195; Conservative 50; Mismatches 71; Indels 12; Gaps 3;  
  
QY 5 IKVACVQAAPYMDLEATYDKTIELMEBAARNARLIAPETWIPGYPMFLMDSPAMM 64  
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:  
DB LKVAAVQAAPAFLLDDASIEKTIKYDEAAAAGAKLIAPEPTWIPGYPMWILGAPAMAI 67  
65 Q--FPROHENSLELDGPOAKRISDAKRLGIMVTIGMSERVGTLTYSQWFIGNDGTI 122  
68 MKGFPSRYFDNSLQYSGPEAERLRDAARHNKIYIALGLSERDGSGLYIQWTLIGPGETV 127  
QY 123 GARRLKPTFFVERTLFGEDGSSLAVFETSVGRIGGLCCMEHLQPLTKYALANQNEIHC 182  
DB 128 AQRRLKPTFHARITYFGEDGSHLAVHDDIGRLGLCCMEHLQPLSKYAMTAQNEQYHV 187  
QY 183 AAMPFSLYPNAKALGPDVNVAAASRIYAVEGQCFVLASCALVSGSMIDMLCTDDEKAL 242  
DB 188 AAMPFSLYDPFAHLGAEVNNAAASKIYAVESCFYIACATVQSWIDELCTDEKIQF 247  
QY 243 LLAGGSHRIIGPDGDLVAPLAENEBGILYANLDPGVRIILAKMAADPAHGYSRDPITRL 302  
DB 248 LHVGGGFAVIYGPDSGPLAKPLPPDQEGILYADIDIGMISVAKAADPAHGYSRDPITRL 307  
QY 303 LIDRSP-----KLPIV---VEIEGDLR 320  
DB 308 LFNRRPANKVEKIALFVDDQEAIVDSPLK 335  
  
RESULT 14  
Q6RWS3\_9Z2ZZZ PRELIMINARY; PRT; 337 AA.  
AC Q6RWS3;  
DT 05-JUL-2004 (TrEMBLrel. 27, Created)  
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)  
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)  
DE Nitrilase (EC 3.5.5.7).  
GN ORFNames=BD7048;  
OS unclutered organism.  
OC unclassified; environmental samples.  
OX NCBI\_Taxid=155900;  
RN [1]\_Nucleotide SEQUENCE.  
RP PubMed=15066841; DOI=10.1128/AEM.70.4.2429-2436.2004;  
RA Robertson D.E., Chaplin J.A., Desantis G., Podar M., Madden M.,  
RA Chi E., Richardson T., Milan A., Miller M., Weiner D.P., Wong K.,  
RA McQuaid J., Farwell B., Preston L.A., Tan X., Sneed M.A., Keller M.,  
RA Mathur E., Kretz P.L., Burk M.J., Short J.M.;  
RT "Exploring nitrilase sequence space for enantioselective catalysis.";  
RL Appl. Environ. Microbiol. 70:2429-2436(2004).  
DR EMBL; AY487429; AAR97376.1; -; Genomic DNA.  
DR GO; GO:0018762; F:aliphatic nitrilase activity; IEA.  
DR GO; GO:0016810; F:hydrolyase activity, acting on carbon-nitrog. .; IEA.  
DR GO; GO:0006807; P:nitrogen compound metabolism; IEA.

DR InterPro; IPR000132; N/CN hydrolase.  
DR InterPro; IPR003010; Ntlase/CNhydase.  
DR Pfam; PF00795; CN hydrolase; 1.  
DR PROSITE; PS50263; CN HYDROLASE; 1.  
DR PROSITE; PS00920; NITRIL\_CHT\_1; 1.  
DR PROSITE; PS00921; NITRIL\_CHT\_2; 1.  
DR Hydrolyase.  
SQ SEQUENCE 337 AA; 36423 MW; 841BABA391031A81 CRC64;  
  
Query Match 58.5%; Score 1033; DB 2; Length 337;  
Best Local Similarity 59.3%; Pred. No. 2.7e-81;  
Matches 192; Conservative 55; Mismatches 67; Indels 10; Gaps 3;  
  
QY 6 KVACVQAAPYMDLEATYDKTIELMEBAARNARLIAPETWIPGYPMFLMDSPAMM 64  
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:  
DB 9 KVAVQAAPAMLDLDDASIAKSTALTEBAAGAKLIAPEAPITPYPMWILGAPAMAI 68  
65 -QFVROHENSLELDGPOAKRISDAKRLGIMVTIGMSERVGTLTYSQWFIGNDGTI 123  
69 RGFVGRYFDNSLAVDSPOAKRLAVKAGLTAIVIGLSEREGSLYLAQWILGPDGETIA 128  
QY 124 ARRLKPTFFVERTLFGEDGSSLAVF--TSVGRIGGLCCMEHLQPLTKYALANQNEIHC 182  
DB 129 KRRRLKPTFHARITYFGEDGSDLAHVRPGIGRLGLCCMEHLQPLSKYAMTAQNEQYHV 188  
QY 183 AAMPFSLYPNAKALGPDVNVAAASRIYAVEGQCFVLASCALVSGSMIDMLCTDDEKAL 242  
DB 189 AAMPFSLYDPFAHLGAEVNNAAASRIYAVESCFVLAPCATVQSWIDELCTDDEKAL 248  
QY 243 LLAGGSHRIIGPDGDLVAPLAENEBGILYANLDPGVRIILAKMAADPAHGYSRDPITRL 302  
DB 249 LHVGGGHAATIGPDSSIAEKLPPDQEGILYADIDIGMISVAKAADPAHGYSRDPITRL 308  
QY 303 LIDRSP-----KLPIV---VEIEGDL 319  
DB 309 LNNKPSKRVHFSLPVDNVEPI 332  
  
RESULT 15  
Q6RWE4\_9Z2ZZZ PRELIMINARY; PRT; 338 AA.  
AC Q6RWE4;  
DT 05-JUL-2004 (TrEMBLrel. 27, Created)  
DT 05-JUL-2004 (TrEMBLrel. 27, Last sequence update)  
DT 05-JUL-2004 (TrEMBLrel. 27, Last annotation update)  
DE Nitrilase (EC 3.5.5.7).  
GN ORFNames=BD5287;  
OS unclutered organism.  
OC unclassified; environmental samples.  
OX NCBI\_Taxid=155900;  
RN [1]\_Nucleotide SEQUENCE.  
RP PubMed=15066841; DOI=10.1128/AEM.70.4.2429-2436.2004;  
RA Robertson D.E., Chaplin J.A., Desantis G., Podar M., Madden M.,  
RA Chi E., Richardson T., Milan A., Miller M., Weiner D.P., Wong K.,  
RA McQuaid J., Farwell B., Preston L.A., Tan X., Sneed M.A., Keller M.,  
RA Mathur E., Kretz P.L., Burk M.J., Short J.M.;  
RT "Exploring nitrilase sequence space for enantioselective catalysis.";  
RL Appl. Environ. Microbiol. 70:2429-2436(2004).  
DR EMBL; AY487558; AAR97505.1; -; Genomic DNA.  
DR GO; GO:0018762; F:aliphatic nitrilase activity; IEA.  
DR GO; GO:0016810; F:hydrolyase activity, acting on carbon-nitrog. .; IEA.  
DR GO; GO:0006807; P:nitrogen compound metabolism; IEA.  
DR InterPro; IPR000132; N/CN hydrolase.  
DR InterPro; IPR003010; Ntlase/CNhydase.  
DR Pfam; PF00795; CN hydrolase; 1.  
DR PROSITE; PS50263; CN HYDROLASE; 1.  
DR PROSITE; PS00920; NITRIL\_CHT\_1; 1.  
DR PROSITE; PS00921; NITRIL\_CHT\_2; 1.  
DR Hydrolyase.  
SQ SEQUENCE 338 AA; 36487 MW; 22F4109903344F59 CRC64;  
  
Query Match 58.5%; Score 1033; DB 2; Length 338;

Best Local Similarity 60.1%; Pred. No. 2.7e-81;  
Matches 194; Conservative 53; Mismatches 66; Indels 10; Gaps 3;

QY	6	KVACVQAPIMYDIEATVDKTIEMEEAARNNAFLIAFPETWIPGYPFWFLMDS PANAM- 64
Db	9	RVAVVQAAPAMLDLEASVSKSIALIEEAAKAKLIAFPFAFIPGYFWYIMLDS PANAMIG 68
QY	65	-QPVROYHENSLELDGPQAKRISDAKRLGIMVTIGMSERVGTLYISQWFIQDNGDTIG 123
Db	69	RGFVQRYFDNSLSYDSPQAEIRLRLAVKKAGMTAVLGLSERDGSILYLAQWMLIGPDGETIA 128
QY	124	ARRLKPTFFVERTLFEEDGSSIAVEE-TSVGRIGLGCWEHLOPLTKYALYAONEBIIHC 182
Db	129	KRRKLRTTHAERTYVGEEDGSDLAHVDRPGIGRLGALCWEHLOPLSKYAMTAQNEQYHV 188
QY	183	AAMPFSILYPNAKALGPDVNVVAASRIYAVEGQCFVLA SCALVSGSMIDMLCTDDEKHAL 242
Db	189	AAMPFSILYDPFAPALGWEVNNAASRYVAVEGSCFVLAPCATVSGAMIDELCDRDKHAL 248
QY	243	LLAGGHSRIIGPDGGDLVAPLAENEGILYANLDPGVRIILAKMAADPAGHSRPDITRL 302
Db	249	LHVGGHAAIFGPPDGSATADKLPSDQEGILPADIDLGAIGAKNADDPAGHSRPDVTRL 308
QY	303	LIDRSPK-----LPVVEIEDG 318
Db	309	LUNKKPSKVEHFALPLDTLAGE 331

Search completed: April 27, 2006, 01:00:11  
Job time : 152.477 secs



GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

## OM protein - protein search, using sw model

Run on: April 27, 2006, 00:50:05 ; Search time 132.234 Seconds  
(without alignments)  
1119.760 Million cell updates/sec

Title: US-09-751-299-4

Perfect score: 1765  
Sequence: 1 MKEAIKVCACVGAAPITMDL.....DLRFVALGKASGTGAQLLEI 337

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 2443163 seqs, 439378781 residues

Total number of hits satisfying chosen parameters: 2443163

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

## Database :

A. Geneseq\_21:\*  
1: geneseqp1980s:\*  
2: geneseqp1990s:\*  
3: geneseqp2000s:\*  
4: geneseqp2001s:\*  
5: geneseqp2002s:\*  
6: geneseqp2003as:\*  
7: geneseqp2003bs:\*  
8: geneseqp2004s:\*  
9: geneseqp2005s:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	length	DB ID	Description
1	1765	100.0	337	4	AAE05489 Nitrlase
2	1761	99.8	337	7	ADC24119 Protein s
3	1761	99.8	337	8	ADH36220 Chemical
4	1761	99.8	337	8	ADG93919 Nitrlase
5	1761	99.8	337	8	Adi62518 Nitrlase
6	1761	99.8	337	8	Adi64639 Nitrlase
7	1339	75.9	337	7	ADC23791 Protein s
8	1339	75.9	337	7	ADC23781 Protein s
9	1339	75.9	337	8	ADH35882 Chemical
10	1339	75.9	337	8	ADH35892 Chemical
11	1339	75.9	337	8	ADG93583 Nitrlase
12	1339	75.9	337	8	ADG93593 Nitrlase
13	1339	75.9	337	8	Adi62180 Nitrlase
14	1339	75.9	337	8	Adi62190 Nitrlase
15	1339	75.9	337	8	Adi64311 Nitrlase
16	1339	75.9	337	8	Adi64301 Nitrlase
17	1339	64.5	333	7	ADC24025 Protein s
18	1139	64.5	333	8	ADH36126 Chemical
19	1139	64.5	333	8	ADG93827 Nitrlase
20	1139	64.5	333	8	Adi62424 Nitrlase
21	1139	64.5	333	8	Adi64545 Nitrlase
22	1136	64.4	332	7	ADG23873 Protein s
23	1136	64.4	332	8	ADH35974 Chemical
24	1136	64.4	332	8	ADG93675 Nitrlase

25	1136	64.4	332	8	Adi62272 Nitrlase
26	1136	64.4	332	8	Adi64393 Nitrlase
27	1091	61.8	355	7	ADC24039 Protein s
28	1091	61.8	355	8	ADH36140 Chemical
29	1091	61.8	355	8	ADG93841 Nitrlase
30	1091	61.8	355	8	Adi62438 Nitrlase
31	1091	61.8	355	8	Adi64559 Nitrlase
32	1081	61.2	345	7	ADC24067 Protein s
33	1081	61.2	345	8	ADH36168 Chemical
34	1081	61.2	345	8	ADG93868 Nitrlase
35	1081	61.2	345	8	Adi62466 Nitrlase
36	1081	61.2	345	8	Adi64587 Nitrlase
37	1051	59.5	334	7	ADC23739 Protein s
38	1051	59.5	334	8	ADH35840 Chemical
39	1051	59.5	334	8	ADG93541 Nitrlase
40	1051	59.5	334	8	Adi62138 Nitrlase
41	1051	59.5	334	8	Adi64259 Nitrlase
42	1042	59.0	353	7	ADC23945 Protein s
43	1042	59.0	353	8	ADH36046 Chemical
44	1042	59.0	353	8	ADG93747 Nitrlase
45	1042	59.0	353	8	Adi62344 Nitrlase

## ALIGNMENTS

RESULT 1	AAE05489	standard; protein; 337 AA.
ID	AAE05489	
AC	AAE05489;	
DT	25-SEP-2001	(first entry)
DE	Nitrlase protein #2.	
KW	Nitrlase; enantiomer; alpha-substituted carboxylic acid; stereoselective; hydrolysis; amino nitrile; cyanohydrin.	
OS	Unidentified.	
PN	WC200148175-A2.	
PP	05-JUL-2001.	
PF	29-DEC-2000; 2000WO-US035555.	
PR	29-DEC-1999; 99US-0173609P.	
XX	07-DEC-2000; 2000US-0254414P.	
XX	(DIVE-) DIVERSA CORP.	
XX	(MAD/) MADSEN D.	
XX	Madden M, Weiner DP, Chaplin JA;	
XX	WPI; 2001-465211/50.	
XX	N-PSDB; AAD11167.	
XX	Producing alpha-substituted carboxylic acid enantiomers by contacting aldehyde or ketone with cyanide, ammonia compound or its salt or amine and hydrolyzing the resulting amino nitrile or cyanohydrin with Nitrlase.	
XX	Claim 30; Page 86-87; 87pp; English.	
XX	The present invention relates to method for producing an enantiomerically pure alpha-substituted carboxylic acid. The method involves contacting an aldehyde or ketone with a cyanide containing compound and an ammonia-containing compound or an ammonium salt or amine, and stereoselectively hydrolysing the resulting amino nitrile or cyanohydrin intermediate with a nitrlase or a polypeptide having nitrlase activity. The present sequence is a nitrlase protein #2	

Sequence 337 AA:  
Query Match 100.0%; Score 1765; DB 4; Length 337;  
Best Local Similarity 100.0%; Pred. No. 7,7e-188;  
Matches 337; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKEAIKVAACVQAAPTYMDLEATVDKTELMEEBAARNRRLIAPETWIPGYPWFLMDSP 60  
DB 1 MKEAIKVAACVQAAPTYMDLEATVDKTELMEEBAARNRRLIAPETWIPGYPWFLMDSP 60  
QY 61 AMAQFVROYHENSLELDGPQAKRISDAKRLGIMVTLGMSERVGTLTYSQWFIGDNGD 120  
DB 61 AMAQFVROYHENSLELDGPQAKRISDAKRLGIMVTLGMSERVGTLTYSQWFIGDNGD 120  
QY 121 TIGARRKIKPTFVERTLFGEDGSSLAVFETSVGRLGGLCWEHLQPLTKYALYAQNEEI 180  
DB 121 TIGARRKIKPTFVERTLFGEDGSSLAVFETSVGRLGGLCWEHLQPLTKYALYAQNEEI 180  
QY 181 HCAAMPSPSLYPNAKALGPDVNVAASRIYAVEGQCFVLASCALVSQSMIDMLCTDDEKH 240  
DB 181 HCAAMPSPSLYPNAKALGPDVNVAASRIYAVEGQCFVLASCALVSQSMIDMLCTDDEKH 240  
QY 241 ALLLAGGSHRIIGPDGDLVAPLAENEGILYANLDPGVRIILAKMAADPAGHSRPDIT 300  
DB 241 ALLLAGGSHRIIGPDGDLVAPLAENEGILYANLDPGVRIILAKMAADPAGHSRPDIT 300  
QY 301 RLILDRSPKLPVVEIEGDLRPYALGKASETGAQLEEI 337  
DB 301 RLILDRSPKLPVVEIEGDLRPYALGKASETGAQLEEI 337

RESULT 2  
ADCC24119  
ID ADCC24119 standard; protein; 337 AA.  
XX  
AC ADCC24119;  
XX  
DT 18-DEC-2003 (first entry)  
XX  
DE Protein sequence (SeqID 386) exhibiting nitrilase activity.  
XX  
KW enzyme; nitrilase; nitrile; cyanohydrin; ammonia; biocatalyst;  
KW enantiomer; chiral medicine.  
XX  
OS Unidentified.  
XX  
PN WO200300840-A2.  
XX  
PD 03-JAN-2003.  
XX  
PF 15-MAY-2002; 2002WO-US015983.  
XX  
PR 21-JUN-2001; 2001US-0300189P.  
PR 30-JUL-2001; 2001US-030906P.  
PR 22-JAN-2002; 2002US-0351336P.  
XX  
PA (DIVE-) DIVERSA CORP.  
PA (MADD/) MADDEN D.  
XX  
PI Madden M, Desantis G, Chaplin JA, Weiner DP, Milan A, Chi E,  
PI Short JM, Burk M;  
XX  
DR WPI; 2003-201417/19.  
DR N-PSDB; ADCC24118.  
XX  
XX Novel nitrilase polypeptide, useful for making (R)- or (S)-ethyl-4-cyano-  
PT 3-hydroxybutyric acid or (R)- or (S)-mandelic acid or (S)- or (R)-phenyl  
PT lactic acid derivative and for producing pharmaceutical composition, and  
PT food additive.  
XX  
PS Claim 40; SEQ ID NO 386; 560pp; English.  
XX  
XX This invention relates to nitrilases and the nucleic acids that encode

these enzymes thereof. Specifically, it refers to polypeptides that exhibit nitrilase activity, i.e. the ability to directly hydrolyse nitriles or cyanohydrins into their corresponding carboxylic acids and ammonia. Nitrilases have commercial utility as biocatalysts for use in the synthesis of enantiomerically pure aromatic and aliphatic amino acids, as well as hydroxy acids, which are important for the development of chiral medicines. Furthermore, the present invention describes nitrilases, isolated from mesophilic microorganisms, that have improved activity and stability at increased pH and temperature. They are also inexpensive, efficient catalysts, have broad substrate specificity and are capable of chiral differentiation. This polypeptide is a protein sequence that exhibits nitrilase activity of the invention.

Sequence 337 AA:  
Query Match 99.8%; Score 1761; DB 7; Length 337;  
Best Local Similarity 99.7%; Pred. No. 2.2e-187;  
Matches 336; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKEAIKVAACVQAAPTYMDLEATVDKTELMEEBAARNRRLIAPETWIPGYPWFLMDSP 60  
DB 1 MKEAIKVAACVQAAPTYMDLEATVDKTELMEEBAARNRRLIAPETWIPGYPWFLMDSP 60  
QY 61 AMAQFVROYHENSLELDGPQAKRISDAKRLGIMVTLGMSERVGTLTYSQWFIGDNGD 120  
DB 61 AMAQFVROYHENSLELDGPQAKRISDAKRLGIMVTLGMSERVGTLTYSQWFIGDNGD 120  
QY 121 TIGARRKIKPTFVERTLFGEDGSSLAVFETSVGRLGGLCWEHLQPLTKYALYAQNEEI 180  
DB 121 TIGARRKIKPTFVERTLFGEDGSSLAVFETSVGRLGGLCWEHLQPLTKYALYAQNEEI 180  
QY 181 HCAAMPSPSLYPNAKALGPDVNVAASRIYAVEGQCFVLASCALVSQSMIDMLCTDDEKH 240  
DB 181 HCAAMPSPSLYPNAKALGPDVNVAASRIYAVEGQCFVLASCALVSQSMIDMLCTDDEKH 240  
QY 241 ALLLAGGSHRIIGPDGDLVAPLAENEGILYANLDPGVRIILAKMAADPAGHSRPDIT 300  
DB 241 ALLLAGGSHRIIGPDGDLVAPLAENEGILYANLDPGVRIILAKMAADPAGHSRPDIT 300  
QY 301 RLILDRSPKLPVVEIEGDLRPYALGKASETGAQLEEI 337  
DB 301 RLILDRSPKLPVVEIEGDLRPYALGKASETGAQLEEI 337

RESULT 3  
ADH36220  
ID ADH36220 standard; protein; 337 AA.  
XX  
AC ADH36220;  
XX  
DT 11-MAR-2004 (first entry)  
XX  
DE Chemical process monitoring-related nitrilase protein sequence SeqID386.  
XX  
KW chemical process monitoring; biochemical process monitoring; cyanide;  
KW high throughput system; enzyme.  
XX  
OS Unidentified.  
XX  
PN WO2003098187-A2.  
XX  
PD 27-NOV-2003.  
XX  
PF 15-MAY-2003; 2003WO-US015639.  
XX  
PR 15-MAY-2002; 2002US-0380737P.  
XX  
PA (DIVE-) DIVERSA CORP.  
XX  
PI Weiner D, Chaplin JA, Chi E, Milan A, Desantis G, Burk MJ;  
PI McQuaid J, Sege U;  
XX  
DR WPI; 2004-142708/14.

DR N-PSDB; ADH36219.  
XX Monitoring a chemical or biochemical process comprises providing a  
PT reactant comprising a cyanide or a material that can be converted to  
PT cyanide or a reactant that generates a cyanide or a material that can be  
PT converted to cyanide.  
XX  
PS Claim 73; SEQ ID NO 386; 277bp; English.  
XX  
CC This invention relates to a novel method of monitoring chemical or  
CC biochemical processes. The method involves providing a reactant  
CC comprising cyanide (or a material that can be converted to a cyanide)  
CC that generates as a reaction product cyanide or a material that can be  
CC converted to cyanide and measuring the concentration of produced cyanide.  
CC The method is useful for monitoring a chemical or biochemical process.  
CC The method is effective for high throughput systems and is sufficiently  
CC sensitive to detect a small amount of product. The present sequence is  
CC that of a nitrilase protein which may be used in the method of the  
CC invention.  
XX  
SQ Sequence 337 AA;  
Query Match 99.8%; Score 1761; DB 8; Length 337;  
Best Local Similarity 99.7%; Pred. No. 2.2e-187;  
Matches 336; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MKEAIKVCVQAAPIYMDLEATVDKTIEMEBARNNAFLAPETWIPGYPMFLMDSP 60  
DB 1 MKEAIKVCVQAAPIYMDLEATVDKTIEMEBARNNAFLAPETWIPGYPMFLMDSP 60  
QY 61 AWAQFVQYHNSLELDGPOAKRISDAKRLGIVWTLGMSRVRGTYIISQWFIGDNGD 120  
DB 61 AWAQFVQYHNSLELDGPOAKRISDAKRLGIVWTLGMSRVRGTYIISQWFIGDNGD 120  
QY 121 TIGARRKIKPTFVERTLFGEGDSSLAFFETSVGRGLGCWEHLQPLTKYALYAQNEEI 180  
DB 121 TIGARRKIKPTFVERTLFGEGDSSLAFFETSVGRGLGCWEHLQPLTKYALYAQNEEI 180  
QY 181 HCAAMPSTSLYPNAKALGPDVNVASRIYAVEGCFVLASCALVSQSMIDMLCTDDEKH 240  
DB 181 HCAAMPSTSLYPNAKALGPDVNVASRIYAVEGCFVLASCALVSQSMIDMLCTDDEKH 240  
QY 241 ALLLAGGSHRIIGPDGDLVAPLAENEGILYANLDPGVRIILAMAADPAGHSRPDIT 300  
DB 241 ALLLAGGSHRIIGPDGDLVAPLAENEGILYANLDPGVRIILAMAADPAGHSRPDIT 300  
QY 301 RLIDRSPKLPVVEIEGDLRPYALGKASSETGAQLEEI 337  
DB 301 RLIDRSPKLPVVEIEGDLRPYALGKASSETGAQLEEI 337  
RESULT 4  
ADG93919  
ID ADG93919 standard; protein; 337 AA.  
XX  
AC ADG93919;  
XX  
DT 11-MAR-2004 (first entry)  
XX  
DE Nitrilase enzyme amino acid sequence SeqID386.  
XX  
KM nitrilase; nitrile; carboxylic acid; chemical process; pH; temperature;  
KM enantioselective transformation; enzyme.  
XX  
OS Unidentified.  
XX  
PN MO2003097810-A2.  
XX  
PD 27-NOV-2003.  
XX  
PF 15-MAY-2003; 2003WO-US015712.  
XX  
PR 15-MAY-2002; 2002US-00146772.

PR 09-SEP-2002; 2002US-00241742.  
XX  
XX (DIVE-) DIVERSA CORP.  
XX  
XX Desantis G, Short JM, Burk M, Wong K, Farwell R, Chatman K;  
XX WPI; 2004-090637/09.  
DR N-PSDB; ADG93918.  
XX  
XX  
PT New isolated or recombinant nucleic acid encoding a polypeptide having  
PT nitrilase activity, useful for screening enantioselective transformation.  
XX  
PS Claim 44; SEQ ID NO 386; 295bp; English.  
XX  
CC This invention is related to a novel isolated or recombinant nucleic acid  
CC encoding a protein having nitrilase activity. Nitrilase's are capable of  
CC converting nitrile's directly to carboxylic acids and have great  
CC potential for use in industrial chemical processes. The isolated  
CC nitrilase proteins of the invention have increased activity and stability  
CC at increased pH and temperature when compared to those conventionally  
CC used. In addition, the nucleic acid of the invention is useful for  
CC screening enantioselective transformation. The present sequence is that  
CC of a nitrilase enzyme of the invention.  
XX  
SQ Sequence 337 AA;  
Query Match 99.8%; Score 1761; DB 8; Length 337;  
Best Local Similarity 99.7%; Pred. No. 2.2e-187;  
Matches 336; Conservative 1; Mismatches 0; Indels 0; Gaps 0;  
QY 1 MKEAIKVCVQAAPIYMDLEATVDKTIEMEBARNNAFLAPETWIPGYPMFLMDSP 60  
DB 1 MKEAIKVCVQAAPIYMDLEATVDKTIEMEBARNNAFLAPETWIPGYPMFLMDSP 60  
QY 61 AWAQFVQYHNSLELDGPOAKRISDAKRLGIVWTLGMSRVRGTYIISQWFIGDNGD 120  
DB 61 AWAQFVQYHNSLELDGPOAKRISDAKRLGIVWTLGMSRVRGTYIISQWFIGDNGD 120  
QY 121 TIGARRKIKPTFVERTLFGEGDSSLAFFETSVGRGLGCWEHLQPLTKYALYAQNEEI 180  
DB 121 TIGARRKIKPTFVERTLFGEGDSSLAFFETSVGRGLGCWEHLQPLTKYALYAQNEEI 180  
QY 181 HCAAMPSTSLYPNAKALGPDVNVASRIYAVEGCFVLASCALVSQSMIDMLCTDDEKH 240  
DB 181 HCAAMPSTSLYPNAKALGPDVNVASRIYAVEGCFVLASCALVSQSMIDMLCTDDEKH 240  
QY 241 ALLLAGGSHRIIGPDGDLVAPLAENEGILYANLDPGVRIILAMAADPAGHSRPDIT 300  
DB 241 ALLLAGGSHRIIGPDGDLVAPLAENEGILYANLDPGVRIILAMAADPAGHSRPDIT 300  
QY 301 RLIDRSPKLPVVEIEGDLRPYALGKASSETGAQLEEI 337  
DB 301 RLIDRSPKLPVVEIEGDLRPYALGKASSETGAQLEEI 337  
RESULT 5  
AD162518  
ID AD162518 standard; protein; 337 AA.  
XX  
AC AD162518;  
XX  
DT 22-APR-2004 (first entry)  
XX  
DE Nitrilase polypeptide #193.  
XX  
KM Acorvaetatin; (R)-ethyl 4-cyano-3-hydroxybutyrate;  
KM (R)-ethyl 4-cyano-3-hydroxybutyric acid; epichlorohydrin;  
KM 3-hydroxyglutaronitrile; 3-hydroxyglutaronitrile;  
KM 4-cyano-3-hydroxybutyric acid; ethyl 4-cyano-3-hydroxybutyric acid;  
KM mixed hyperlipidaemia; homozygous familial hypercholesterolaemia;  
KM antilipemic; enzyme.  
XX  
OS Unidentified.

```

XX  W02003106415-A2.
PN
XX
XX  24-DEC-2003.
PD
XX
XX  13-JUN-2003; 2003WO-US018840.
PF
XX
XX  13-JUN-2002; 2002US-0389317P.
PR
XX  28-JUN-2002; 2002US-0392944P.
PR
XX  (DIVE-) DIVERSA CORP.
PA
XX  (DIVE-) DIVERSA CORP.
PI
XX  Burk M, Desantis G, Morgan B, Zhu Z;
XX  MPI; 2004-090821/09.
XX  N-PSDB; ADI62517.
DR
XX
XX  Preparation of atorvastatin comprises catalytic conversion of 3-
PT  hydroxyglutaronitrile by polypeptide with nitrilase activity, converting
PT  obtained 4-cyano-3-hydroxybutyric acid to ethyl-4-cyano-3-hydroxybutyric
PT  acid and forming atorvastatin.
XX
XX  Claim 41; SEQ ID NO 386; 253bp; English.
PS
XX
XX  The present invention relates to a method for preparing an atorvastatin
CC  intermediate known as (R)-ethyl 4-cyano-3-hydroxybutyrate ((R)-ethyl 4-
CC  cyano-3-hydroxybutyric acid). The method comprises optionally converting
CC  epichlorohydrin or equivalent to 3-hydroxyglutaronitrile, catalytic
CC  conversion of 3-hydroxyglutaronitrile or equivalent to 4-cyano-3-
CC  hydroxybutyric acid with a polypeptide having nitrilase activity,
CC  converting 4-cyano-3-hydroxybutyric to ethyl-4-cyano-3-hydroxybutyric
CC  acid, and converting this to (R)-ethyl 4-cyano-3-hydroxybutyrate. The
CC  method involves whole cell processes, cell lysate process, "one pot"
CC  processes, and "multi-pot" processes using a variety of parameters.
CC  Atorvastatin is used, in conjunction with dietary restriction, in the
CC  management of hyperlipidaemia, including hypercholesterolaemia, mixed
CC  dyslipidaemia and homozygous familial hypercholesterolaemia. The present
CC  sequence represents a nitrilase polypeptide obtained from an
CC  environmental sample.
XX
XX  Sequence 337 AA;
SQ
XX
XX  Query Match          99.8%; Score 1761; DB 8; Length 337;
XX  Best Local Similarity 99.7%; Pred. No. 2,2e-187;
XX  Matches 336; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
XX
XX  1 MKEAIRVACVQAAPIYMDIEATVDTKTIELMEBAARNNAFLAEPETWIPGYWFLMLDSP 60
XX  1 MKEAIRVACVQAAPIYMDIEATVDTKTIELMEBAARNNAFLAEPETWIPGYWFLMLDSP 60
XX
XX  61 AAMNQVROVHENSLELDGPOAKRISDAKRLGIMWTLGMSERVGGLYISQWFIGDND 120
XX  61 AAMNQVROVHENSLELDGPOAKRISDAKRLGIMWTLGMSERVGGLYISQWFIGDND 120
XX
XX  121 TIGARRKLTKEFTFERTLFGEGDSSLAIVETSVGRIGLGCWEHLPLTKYALYAQNEEI 180
XX  121 TIGARRKLTKEFTFERTLFGEGDSSLAIVETSVGRIGLGCWEHLPLTKYALYAQNEEI 180
XX
XX  122 TIGARRKLTKEFTFERTLFGEGDSSLAIVETSVGRIGLGCWEHLPLTKYALYAQNEEI 180
XX  122 TIGARRKLTKEFTFERTLFGEGDSSLAIVETSVGRIGLGCWEHLPLTKYALYAQNEEI 180
XX
XX  181 HCAAMPFSFSLYPNAAKALGPDVNAASRIYAVGQCFLVASCALVSQSMIDMLCTDDEK 240
XX  181 HCAAMPFSFSLYPNAAKALGPDVNAASRIYAVGQCFLVASCALVSQSMIDMLCTDDEK 240
XX
XX  181 HCAAMPFSFSLYPNAAKALGPDVNAASRIYAVGQCFLVASCALVSQSMIDMLCTDDEK 240
XX  181 HCAAMPFSFSLYPNAAKALGPDVNAASRIYAVGQCFLVASCALVSQSMIDMLCTDDEK 240
XX
XX  241 ALLAGGSHRITIGPDGSDIVAPLAENEGILLVANIADPGVRIILAKMAADPAGHSRPDIT 300
XX  241 ALLAGGSHRITIGPDGSDIVAPLAENEGILLVANIADPGVRIILAKMAADPAGHSRPDIT 300
XX
XX  301 RLILDRSPKLPVVEIEGDLRPYALGKASEGQOLEEI 337
XX  301 RLILDRSPKLPVVEIEGDLRPYALGKASEGQOLEEI 337
XX
XX  301 RLILDRSPKLPVVEIEGDLRPYALGKASEGQOLEEI 337
XX  301 RLILDRSPKLPVVEIEGDLRPYALGKASEGQOLEEI 337
XX
XX  RESULT 6
XX  ADI64639

```

```

ID  ADI64639 standard; protein; 337 AA.
XX
XX  ADI64639;
AC
XX
XX  22-APR-2004 (first entry)
DT
XX
XX  Nitrilase seq id 194.
DE
XX
XX  (R)-ethyl 4-cyano-3-hydroxybutyric acid; nitrile hydrolysis;
XX  carboxylic acid; cyanohydrin moiety hydrolysis;
XX  ammonitrile moiety hydrolysis; chiral alpha-hydroxy acid molecule;
XX  chiral amino acid molecule; (S)-ethyl 4-cyano-3-hydroxybutyric acid;
XX  (R)-mandelic acid; (S)-mandelic acid; (S)-phenyl lactic acid derivative;
XX  (R)-phenyl lactic acid derivative; & enantiomeric excess;
XX  & diastomeric excess; food additive; drug intermediate; nitrilase.
XX
XX  Unidentified.
OS
XX
XX  US2004014195-A1.
XX
XX  22-JAN-2004.
XX
XX
XX  15-MAY-2003; 2003US-00440523.
XX
XX  29-DEC-1999; 99US-0173609P.
XX  07-DEC-2000; 2000US-0254414P.
XX  28-DEC-2000; 2000US-00751299.
XX  21-JUN-2001; 2001US-0300189P.
XX  30-JUL-2001; 2001US-0309006P.
XX  22-JAN-2002; 2002US-0351316P.
XX  15-MAY-2002; 2002US-00146772.
XX  09-SEP-2002; 2002US-00241742.
XX
XX  (DIVE-) DIVERSA CORP.
PA
XX
XX  Desantis G, Short JM, Burk MJ, Wong K, Farwell R, Chatman K;
XX  MPI; 2004-121569/12.
XX  N-PSDB; ADI64638.
DR
XX
XX  Novel isolated or recombinant polypeptide having nitrilase activity,
PT  useful in production of food additives.
PT
XX
XX  Claim 46; SEQ ID NO 386; 105bp; English.
PS
XX
XX  The invention describes an isolated or recombinant polypeptide (I)
CC  comprising amino acids having a sequence at least 50 % identical to a
CC  sequence (S1) available in electronic form (EC) from the following web
CC  site ftp.segdata.uspro.gov/sequence.html?docid=2004014195, or its
CC  variants, having one or more mutations at residue 55 Lys, Gly or Glu, at
CC  residue 60 glutamic acid, at residue 111 Ser, their combinations or
CC  fragments. (I) is useful for: producing an (R)-ethyl 4-cyano-3-
CC  hydroxybutyric acid; hydrolysing a nitrile to a carboxylic acid;
CC  hydrolysing cyanohydrin moiety or an ammonitrile moiety; producing a
CC  chiral alpha-hydroxy acid molecule or a chiral amino acid molecule;
CC  producing an (S)-ethyl 4-cyano-3-hydroxybutyric acid; producing an (R)-
CC  mandelic acid or (S)-mandelic acid; producing (S)-phenyl lactic acid
CC  derivative or an (R)-phenyl lactic acid derivative; modifying a molecule;
CC  and for identifying a modified compound. The inventive method is useful
CC  for monitoring or determining & enantiomeric excess or & diastomeric
CC  excess. (I) is useful in the production of food additives and drug
CC  intermediates. This is the amino acid sequence of a nitrilase of the
CC  invention.
XX
XX  Sequence 337 AA;
SQ
XX
XX  Query Match          99.8%; Score 1761; DB 8; Length 337;
XX  Best Local Similarity 99.7%; Pred. No. 2,2e-187;
XX  Matches 336; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
XX
XX  1 MKEAIRVACVQAAPIYMDIEATVDTKTIELMEBAARNNAFLAEPETWIPGYWFLMLDSP 60
XX  1 MKEAIRVACVQAAPIYMDIEATVDTKTIELMEBAARNNAFLAEPETWIPGYWFLMLDSP 60
XX
XX  61 AAMNQVROVHENSLELDGPOAKRISDAKRLGIMWTLGMSERVGGLYISQWFIGDND 120
XX  61 AAMNQVROVHENSLELDGPOAKRISDAKRLGIMWTLGMSERVGGLYISQWFIGDND 120
XX
XX  121 TIGARRKLTKEFTFERTLFGEGDSSLAIVETSVGRIGLGCWEHLPLTKYALYAQNEEI 180
XX  121 TIGARRKLTKEFTFERTLFGEGDSSLAIVETSVGRIGLGCWEHLPLTKYALYAQNEEI 180
XX
XX  122 TIGARRKLTKEFTFERTLFGEGDSSLAIVETSVGRIGLGCWEHLPLTKYALYAQNEEI 180
XX  122 TIGARRKLTKEFTFERTLFGEGDSSLAIVETSVGRIGLGCWEHLPLTKYALYAQNEEI 180
XX
XX  181 HCAAMPFSFSLYPNAAKALGPDVNAASRIYAVGQCFLVASCALVSQSMIDMLCTDDEK 240
XX  181 HCAAMPFSFSLYPNAAKALGPDVNAASRIYAVGQCFLVASCALVSQSMIDMLCTDDEK 240
XX
XX  181 HCAAMPFSFSLYPNAAKALGPDVNAASRIYAVGQCFLVASCALVSQSMIDMLCTDDEK 240
XX  181 HCAAMPFSFSLYPNAAKALGPDVNAASRIYAVGQCFLVASCALVSQSMIDMLCTDDEK 240
XX
XX  241 ALLAGGSHRITIGPDGSDIVAPLAENEGILLVANIADPGVRIILAKMAADPAGHSRPDIT 300
XX  241 ALLAGGSHRITIGPDGSDIVAPLAENEGILLVANIADPGVRIILAKMAADPAGHSRPDIT 300
XX
XX  301 RLILDRSPKLPVVEIEGDLRPYALGKASEGQOLEEI 337
XX  301 RLILDRSPKLPVVEIEGDLRPYALGKASEGQOLEEI 337
XX
XX  301 RLILDRSPKLPVVEIEGDLRPYALGKASEGQOLEEI 337
XX  301 RLILDRSPKLPVVEIEGDLRPYALGKASEGQOLEEI 337
XX
XX  RESULT 6
XX  ADI64639

```

QY 61 AAMQFVROYHNSLELDGPOAKRISDAKRLGIMVTLGMSRVGGLYISQWFGDNGD 120  
 DB 61 AAMQFVROYHNSLELDGPOAKRISDAKRLGIMVTLGMSRVGGLYISQWFGDNGD 120  
 QY 121 TIGARRKLPKPTVERTLFEGSDSSLAIVETSVGRGLGCWEHLQPLTKYALYAQNEEI 180  
 DB 121 TIGARRKLPKPTVERTLFEGSDSSLAIVETSVGRGLGCWEHLQPLTKYALYAQNEEI 180  
 QY 181 HCAAMPFSFLYPNAKALGPDVNVAAASRIYAVEGQCFVLASCALVSQSMIDMLCTDDEKH 240  
 DB 181 HCAAMPFSFLYPNAKALGPDVNVAAASRIYAVEGQCFVLASCALVSQSMIDMLCTDDEKH 240  
 QY 241 ALLLAGGSHSRITIGPDGDLVAPLAENEGGILYANLDPGVRIILAKMAADPAGHYSRPDTT 300  
 DB 241 ALLLAGGSHSRITIGPDGDLVAPLAENEGGILYANLDPGVRIILAKMAADPAGHYSRPDTT 300  
 QY 301 RLILDRSPKLPVVEIEGDLRPYALGKASSETGAQLEEI 337  
 DB 301 RLILDRSPKLPVVEIEGDLRPYALGKASSETGAQLEEI 337

RESULT 7  
 ADC23791  
 ID ADC23791 standard; protein; 337 AA.  
 AC ADC23791;  
 DT 18-DEC-2003 (first entry)  
 DE Protein sequence (SeqID 58) exhibiting nitrilase activity.  
 KM enzyme; nitrilase; nitrile; cyanohydrin; ammonia; biocatalyst;  
 KW enantiomer; chiral medicine.  
 OS Unidentified.  
 PN WO2003000840-A2.  
 PD 03-JAN-2003.  
 PF 15-MAY-2002; 2002MO-US015983.  
 PR 21-JUN-2001; 2001US-0300189P.  
 PR 30-JUL-2001; 2001US-0309006P.  
 PR 22-JAN-2002; 2002US-0351336P.  
 PA (DIVE-) DIVERSA CORP.  
 PA (MADD/) MADDEN D.  
 PI Madden M, Desantis G, Chaplin JA, Weiner DP, Milan A, Chi E,  
 PI Short JM, Burk M;  
 DR MPI: 2003-201417/19.  
 DR N-PSDB; ADC23790.  
 XX Novel nitrilase polypeptide, useful for making (R)- or (S)-ethyl-4-cyano-  
 PT 3-hydroxybutyric acid or (R)- or (S)-mandelic acid or (S)- or (R)-phenyl  
 PT lactic acid derivative and for producing pharmaceutical composition, and  
 PT food additive.  
 PS Claim 40; SEQ ID NO 58; 560pp; English.

CC inexpensive, efficient catalysts, have broad substrate specificity and  
 CC are capable of chiral differentiation. This polypeptide is a protein  
 CC sequence that exhibits nitrilase activity of the invention.  
 SQ Sequence 337 AA;  
 Query Match 75.9%; Score 1339; DB 7; Length 337;  
 Best Local Similarity 75.2%; Pred. No. 3.3e-140;  
 Matches 252; Conservative 30; Mismatches 47; Indels 6; Gaps 1;  
 QY 1 MKEAIKVAQVQAPRYMDLEATVDKTIEMERANNNRLAFPEWTIPGYPMFLWDSF 60  
 DB 1 VKEAIKVAQVQAPRYMDLEATVDKTIEMERANNNRLAFPEWTIPGYPMFLWDSF 60  
 QY 61 AAMQFVROYHNSLELDGPOAKRISDAKRLGIMVTLGMSRVGGLYISQWFGDNGD 120  
 DB 61 AAMQFVROYHNSLELDGPOAKRISDAKRLGIMVTLGMSRVGGLYISQWFGDNGD 120  
 QY 121 TIGARRKLPKPTVERTLFEGSDSSLAIVETSVGRGLGCWEHLQPLTKYALYAQNEEI 180  
 DB 121 TIGARRKLPKPTVERTLFEGSDSSLAIVETSVGRGLGCWEHLQPLTKYALYAQNEEI 180  
 QY 181 HCAAMPFSFLYPNAKALGPDVNVAAASRIYAVEGQCFVLASCALVSQSMIDMLCTDDEKH 240  
 DB 181 HCAAMPFSFLYPNAKALGPDVNVAAASRIYAVEGQCFVLASCALVSQSMIDMLCTDDEKH 240  
 QY 241 ALLLAGGSHSRITIGPDGDLVAPLAENEGGILYANLDPGVRIILAKMAADPAGHYSRPDTT 300  
 DB 241 ALLLAGGSHSRITIGPDGDLVAPLAENEGGILYANLDPGVRIILAKMAADPAGHYSRPDTT 300  
 QY 301 RLILDRSPKLPVVEIEGDLRPYALGKASSETGAQLEEI 335  
 DB 301 RLILDRSPKLPVVEIEGDLRPYALGKASSETGAQLEEI 335

RESULT 8  
 ADC23781  
 ID ADC23781 standard; protein; 337 AA.  
 AC ADC23781;  
 DT 18-DEC-2003 (first entry)  
 DE Protein sequence (SeqID 48) exhibiting nitrilase activity.  
 KM enzyme; nitrilase; nitrile; cyanohydrin; ammonia; biocatalyst;  
 KW enantiomer; chiral medicine.  
 OS Unidentified.  
 PN WO2003000840-A2.  
 PD 03-JAN-2003.  
 PF 15-MAY-2002; 2002MO-US015983.  
 PR 21-JUN-2001; 2001US-0300189P.  
 PR 30-JUL-2001; 2001US-0309006P.  
 PR 22-JAN-2002; 2002US-0351336P.  
 PA (DIVE-) DIVERSA CORP.  
 PA (MADD/) MADDEN D.  
 PI Madden M, Desantis G, Chaplin JA, Weiner DP, Milan A, Chi E,  
 PI Short JM, Burk M;  
 DR MPI: 2003-201417/19.  
 DR N-PSDB; ADC23780.  
 XX Novel nitrilase polypeptide, useful for making (R)- or (S)-ethyl-4-cyano-  
 PT 3-hydroxybutyric acid or (R)- or (S)-mandelic acid or (S)- or (R)-phenyl  
 PT lactic acid derivative and for producing pharmaceutical composition, and  
 PT food additive.



XX		enantioselective transformation; enzyme.
XX	.OS	Unidentified.
PN	WO2003097810-A2.	
PD	27-NOV-2003.	
XX		
XX		
PF	15-MAY-2003; 2003WO-US015712.	
XX		
PR	15-MAY-2002; 2002US-00146772.	
XX	09-SEP-2002; 2002US-00241742.	
PA	(DIVE-) DIVERSA CORP.	
PI	Desantis G, Short JM, Burk M, Wong K, Farwell R, Chatman K;	
DR	WPI; 2004-090637/09.	
DR	N-PsDB; ADG93582.	
PT	New isolated or recombinant nucleic acid encoding a polypeptide having	
XX	nitrilase activity, useful for screening enantioselective transformation.	
PS	Claim 44; SEQ ID NO 48; 295dp; English.	
CC	This invention is related to a novel isolated or recombinant nucleic acid	
CC	encoding a protein having nitrilase activity. Nitrilase's are capable of	
CC	converting nitrile's directly to carboxylic acids and have great	
CC	potential for use in industrial chemical processes. The isolated	
CC	nitrilase proteins of the invention have increased activity and stability	
CC	at increased pH and temperature when compared to those conventionally	
CC	used. In addition, the nucleic acid of the invention is useful for	
CC	screening enantioselective transformation. The present sequence is that	
XX	of a nitrilase enzyme of the invention.	
SQ	Sequence 337 AA:	
Query Match	75.9%; Score 1339; DB 8; Length 337;	
Best Local Similarity	75.2%; Pred. No. 3.3e-140;	
Matches 252; Conservative 30; Mismatches 47; Indels 6; Gaps 1		
QY	1 MKEAIKACVQAATITMDLEATVDKTIEMEEAARNRLLTAFPEFTWIPGYWFMILDSP	60
DB	1 VKEAIKACVQAAPVFPLDLDTVDKTYALLIEAARNGRLTAFPEFTWIPGYWFMILDSP	60
QY	61 AMAAQFROYENSLIEDGPQAKRISDAKKRLGITVTIGMSERVGCTLYISOMFIGNDG	120
DB	61 AMGQGFRRRIHNSLVLDSPQAKISEANQAGIYVALGYEERSVGTILYMGMILDDKE	120
QY	121 TTAGRRKLKPTFVRRTLFGEGDGSSLAFFETSVGRLGGLCMEHLOPLTKYALVAQNBER	180
DB	121 TAGRRKLKPTHVERTLFGEEDGSSLSTFPDPLGVLGGLCCMEHLQPLSKYALVAQNBER	180
QY	181 HCAAMPESFLYPNAKAALGPVVNVAASRIYAVEGCFFLASCALVSQMIDMLCTDDEKH	240
DB	181 HFAAMPFSFIYRQATEVLGPEVNVAASRIYAVEGCFFLASCALVSPMIEIMLCIDESKH	240
QY	241 ALLLAGGHSRTIGPPDGDGLVAPLAENNEGILYANLDPGVTLIAQMAADPGHYSRPDT	300
DB	241 SLLOAGGGYSITIGPIDSDLARLPGENNEBGLLYATLTDPAAARIYAKTAADDPAGHYRPDV	300
QY	301 RLILDRSPKLPVVEIEGDLRPYALGKASETGAOLE	335
DB	301 RLILNRSAHQPVVEVGREI-----PASAQFEVE	329
RESULT 12		
ADG93593	ID	ADG93593 standard; protein; 337 AA.
AC	ADG93593;	
XT	11-MAR-2004 (first entry)	

```
XX Nitrilase enzyme amino acid sequence SeqID58.
DE nitrilase; nitrile; carboxylic acid; chemical process; pH; temperature;
KM enantioselective transformation; enzyme.
XX
XX Unidentified.
XX
XX MO2003097810-A2.
XX
XX 27-NOV-2003.
XX
XX 15-MAY-2003; 2003WO-US015712.
XX
XX 15-MAY-2002; 2002US-00146772.
XX
XX 09-SEP-2002; 2002US-00241742.
XX
XX (DIVE-) DIVERSA CORP.
XX
XX Desantis G, Short JM, Burk M, Wong K, Farwell R, Chatman K;
XX
XX WPI; 2004-090637/09.
XX
XX N-PSDB; ADG93592.
XX
XX
XX New isolated or recombinant nucleic acid encoding a polypeptide having
XX nitrilase activity, useful for screening enantioselective transformation.
XX
XX Claim 44; SEQ ID NO 58; 295bp; English.
XX
XX This invention is related to a novel isolated or recombinant nucleic acid
XX encoding a protein having nitrilase activity. Nitrilase's are capable of
XX converting nitrile's directly to carboxylic acids and have great
XX potential for use in industrial chemical processes. The isolated
XX nitrilase proteins of the invention have increased activity and stability
XX at increased pH and temperature when compared to those conventionally
XX used. In addition, the nucleic acid of the invention is useful for
XX screening enantioselective transformation. The present sequence is that
XX of a nitrilase enzyme of the invention.
XX
XX
XX Sequence 337 AA;
XX
XX Query Match 75.9%; Score 1339; DB 8; Length 337;
XX Best Local Similarity 75.2%; Pred. No. 3.3e-140;
XX Matches 252; Conservative 30; Mismatches 47; Indels 6; Gaps 1;
XX
XX 1 MKEAIKVCVQAPAPYMDLEATVDTTIELMEBAARNNAFLIAPETWIPGYPMFLMDSP 60
XX :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX 1 VKEAIKVCVQAPAPFLDLATVDTKTVLIEBAARNGARLIAPETWIPGYPMFLMDSP 60
XX
XX 61 AWAQMPFQRYHENSLELDGPQAKRISDAKRLGIMWTLGMSERVGCTLYISQWFTGDNCD 120
XX :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX 61 AWAQMPFVRRYHENSIVLDSPPQAKRISDAQRAIGIYVALGYSRVSGTLVWGMLDDKGE 120
XX
XX 121 TIGARRKRLKPTFVERTLFGEGDSSLAIVFETSVGRIGLCCMEHLQPLTKYALYAQNEBI 180
XX :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX 121 TAGLRKRLKPTFVERTLFGEGDSSLSLSTFTPLGVLGGLCCMEHLQPLSKYALYAQNEBI 180
XX
XX 181 HCAAMPFSFLYPNAKALGPDVNVAASRIYAVBGCFTVASCALYSQSMIDMLCTDDDEH 240
XX :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX 181 HFAAMPFSFIYQATFVLCPPEVNVASRIYAVBGCFTVASCALYSPEMIEMLCIDESKH 240
XX
XX 241 ALLLGGGGSRIIIGPGGDLVAPLAEENEGIIYANLDPGVRIIAKKAADPAGHSRPDIT 300
XX :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX 241 SLLOAGGGSRIIIGPGGDLARPLGENEGIIYATIDPAARIYAKTAADPAGHSRPDVT 300
XX
XX 301 RLILDSRPLPVVEIEGDLRPYALGASSETGAQLE 335
XX :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX 301 RLILNSANQPVVEVGREI-----PASAQFEVE 329
XX
XX
XX RESULT 13
XX ADI62180
XX ID ADI62180 standard; protein; 337 AA.
```

```
XX ADI62180;
XX
XX 22-APR-2004 (first entry)
XX
XX Nitrilase polypeptide #24.
XX
XX
XX Acorvastatin; (R)-ethyl 4-cyano-3-hydroxybutyrate;
XX (R)-ethyl 4-cyano-3-hydroxybutyric acid; epichlorohydrin;
XX 3-hydroxyglutaronitrile; 3-hydroxyglutaronitrile;
XX 4-cyano-3-hydroxybutyric acid; ethyl-4-cyano-3-hydroxybutyric acid;
XX mixed hyperlipidaemia; homozygous familial hypercholesterolaemia;
XX antilipaseic; enzyme.
XX
XX Unidentified.
XX
XX MO2003106415-A2.
XX
XX 24-DEC-2003.
XX
XX 13-JUN-2003; 2003WO-US018840.
XX
XX 13-JUN-2002; 2002US-0389317P.
XX
XX 28-JUN-2002; 2002US-0392944P.
XX
XX (DIVE-) DIVERSA CORP.
XX
XX Burk M, Desantis G, Morgan B, Zhu Z;
XX
XX WPI; 2004-090821/09.
XX
XX N-PSDB; ADI62179.
XX
XX
XX Preparation of atorvastatin comprises catalytic conversion of 3-
XX hydroxyglutaronitrile by polypeptide with nitrilase activity, converting
XX obtained 4-cyano-3-hydroxybutyric acid to ethyl-4-cyano-3-hydroxybutyric
XX acid and forming atorvastatin.
XX
XX Claim 41; SEQ ID NO 48; 253bp; English.
XX
XX
XX The present invention relates to a method for preparing an atorvastatin
XX intermediate known as (R)-ethyl 4-cyano-3-hydroxybutyrate ((R)-ethyl 4-
XX cyano-3-hydroxybutyric acid). The method comprises optionally converting
XX epichlorohydrin or equivalent to 3-hydroxyglutaronitrile, catalytic
XX conversion of 3-hydroxyglutaronitrile or equivalent to 4-cyano-3-
XX hydroxybutyric acid with a polypeptide having nitrilase activity,
XX converting 4-cyano-3-hydroxybutyric to ethyl-4-cyano-3-hydroxybutyric
XX acid, and converting this to (R)-ethyl 4-cyano-3-hydroxybutyrate. The
XX method involves whole cell processes, cell lysate process, "one pot"
XX processes, and "multi-pot" processes using a variety of parameters.
XX Atorvastatin is used, in conjunction with dietary restriction, in the
XX management of hyperlipidaemia, including hypercholesterolaemia, mixed
XX dyslipidaemia and homozygous familial hypercholesterolaemia. The present
XX sequence represents a nitrilase polypeptide obtained from an
XX environmental sample.
XX
XX
XX Sequence 337 AA;
XX
XX Query Match 75.9%; Score 1339; DB 8; Length 337;
XX Best Local Similarity 75.2%; Pred. No. 3.3e-140;
XX Matches 252; Conservative 30; Mismatches 47; Indels 6; Gaps 1;
XX
XX 1 MKEAIKVCVQAPAPYMDLEATVDTTIELMEBAARNNAFLIAPETWIPGYPMFLMDSP 60
XX :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX 1 VKEAIKVCVQAPAPFLDLATVDTKTVLIEBAARNGARLIAPETWIPGYPMFLMDSP 60
XX
XX 61 AWAQMPFQRYHENSLELDGPQAKRISDAKRLGIMWTLGMSERVGCTLYISQWFTGDNCD 120
XX :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX 61 AWAQMPFVRRYHENSIVLDSPPQAKRISDAQRAIGIYVALGYSRVSGTLVWGMLDDKGE 120
XX
XX 121 TIGARRKRLKPTFVERTLFGEGDSSLAIVFETSVGRIGLCCMEHLQPLTKYALYAQNEBI 180
XX :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
XX 121 TAGLRKRLKPTFVERTLFGEGDSSLSLSTFTPLGVLGGLCCMEHLQPLSKYALYAQNEBI 180
XX
```



QY 181 HCAWPSFSLYPNAKALGPDVNVAAASRIYAVGQCFVLASCALVSQSMIDMLCTDDEKH 240  
 DB 181 HFAWPSFSIYRGATEVILGPEVNVAAASRIYAVGQCFVLASCALVSPEMIEMLCTDDESKH 240  
 QY 241 ALLAGGHSRIITGPDGDLVAPLAENEGILLYANLDPGVRIILAKMAADPAGHSRPDIT 300  
 DB 241 SLQAGGYSRIITGPDGDLVAPLAENEGILLYATLDPARIVAKTAADPAGHSRPDVT 300  
 QY 301 RLIDRSPKLPVVEIRGDLRPVALGKASENGALE 335  
 DB 301 RLINRSANQPVVEVGREI-----PASAGCFEVE 329

RESULT 14  
 ADI62190  
 ID ADI62190 standard; protein; 337 AA.

AC ADI62190;  
 DT 22-APR-2004 (first entry)  
 DE Nitriase polypeptide #29.

XX Acorvastatin; (R)-ethyl 4-cyano-3-hydroxybutyrate;  
 KM (R)-ethyl 4-cyano-3-hydroxybutyric acid; epichlorohydrin;  
 KM 3-hydroxyglutaronitrile; 3-hydroxyglutaronitrile;  
 KM 4-cyano-3-hydroxybutyric acid; ethyl-4-cyano-3-hydroxybutyric acid;  
 KM mixed hyperlipidaemia; homozygous familial hypercholesterolaemia;  
 KM antilipemic; enzyme.

XX Unidentified.  
 XX MO2003106415-A2.

PD 24-DEC-2003.

PF 13-JUN-2003; 2003MO-US018840.

PR 13-JUN-2002; 2002US-0389317P.  
 PR 28-JUN-2002; 2002US-0392944P.

PA (DIVE-) DIVERSA CORP.

PI Burk M, Desantis G, Morgan B, Zhu Z;

XX WPI; 2004-090821/09.

DR N-PSDB; ADI62189.

XX Preparation of atorvastatin comprises catalytic conversion of 3-  
 PT hydroxyglutaronitrile by polypeptide with nitriase activity, converting  
 PT obtained 4-cyano-3-hydroxybutyric acid to ethyl-4-cyano-3-hydroxybutyric  
 PT acid and forming atorvastatin.

XX Claim 41; SEQ ID NO 58; 253pp; English.

XX The present invention relates to a method for preparing an atorvastatin  
 CC intermediate known as (R)-ethyl 4-cyano-3-hydroxybutyrate ((R)-ethyl 4-  
 CC cyano-3-hydroxybutyric acid). The method comprises optionally converting  
 CC epichlorohydrin or equivalent to 3-hydroxyglutaronitrile, catalytic  
 CC conversion of 3-hydroxyglutaronitrile or equivalent to 4-cyano-3-  
 CC hydroxybutyric acid with a polypeptide having nitriase activity,  
 CC converting 4-cyano-3-hydroxybutyric acid to ethyl-4-cyano-3-hydroxybutyric  
 CC acid, and converting this to (R)-ethyl 4-cyano-3-hydroxybutyrate. The  
 CC method involves whole cell processes, cell lysate processes, "one pot"  
 CC processes, and "multi-pot" processes using a variety of parameters.  
 CC Atorvastatin is used, in conjunction with dietary restriction, in the  
 CC management of hyperlipidaemia, including hypercholesterolaemia, mixed  
 CC dyslipidaemia and homozygous familial hypercholesterolaemia. The present  
 CC sequence represents a nitriase polypeptide obtained from an  
 CC environmental sample.

XX Sequence 337 AA;  
 PT

Query Match 75.9%; Score 1339; DB 8; Length 337;  
 Best Local Similarity 75.2%; Pred. No.3.3e-140;  
 Matches 252; Conservative 30; Mismatches 47; Indels 6; Gaps 1;

QY 1 MKEAIRVACVQAAPITMYDLBATYDKTIELMEBARNNARLIAPETWIPGYPMFLWLDSP 60  
 DB 1 VKSAIKVACVQAAPVFLDLATATDKTVALIEBARNGARLIAPETWIPGYPMFLWLDSP 60  
 QY 61 AAMQAFVRCVHENSLELDQARISDAKRLGIMTWLGMSEVGGTLYISQFIGNCD 120  
 DB 61 AAMQAFVRCVHENSLELDQARISDAKRLGIMTWLGMSEVGGTLYISQFIGNCD 120  
 QY 121 TIGARRKLEKPTFERTLFGEGDSSLAVPETSVRGLGCTCHMLDPLTYALYAONEEI 180  
 DB 121 TAGLRKLEKPTFERTLFGEGDSSLAVPETSVRGLGCTCHMLDPLTYALYAONEEI 180  
 QY 181 HCAWPSFSLYPNAKALGPDVNVAAASRIYAVGQCFVLASCALVSQSMIDMLCTDDEKH 240  
 DB 181 HFAWPSFSIYRGATEVILGPEVNVAAASRIYAVGQCFVLASCALVSPEMIEMLCTDDESKH 240  
 QY 241 ALLAGGHSRIITGPDGDLVAPLAENEGILLYANLDPGVRIILAKMAADPAGHSRPDIT 300  
 DB 241 SLQAGGYSRIITGPDGDLVAPLAENEGILLYATLDPARIVAKTAADPAGHSRPDVT 300  
 QY 301 RLIDRSPKLPVVEIRGDLRPVALGKASENGALE 335  
 DB 301 RLINRSANQPVVEVGREI-----PASAGCFEVE 329

RESULT 15  
 ADI64311  
 ID ADI64311 standard; protein; 337 AA.

AC ADI64311;

DT 22-APR-2004 (first entry)

DE Nitriase seq id 30.

XX (R)-ethyl 4-cyano-3-hydroxybutyric acid; nitrile hydrolysis;  
 KM carboxylic acid; cyanohydrin moiety hydrolysis;  
 KM aminonitrile moiety hydrolysis; chiral alpha-hydroxy acid molecule;  
 KM chiral amino acid molecule; (S)-ethyl 4-cyano-3-hydroxybutyric acid;  
 KM (R)-mandelic acid; (S)-mandelic acid; (S)-phenyl lactic acid derivative;  
 KM (R)-phenyl lactic acid derivative; & enantiomeric excess;  
 KM & diastereomeric excess; food additive; drug intermediate; nitriase.

XX Unidentified.

PN US2004014195-A1.

PD 22-JAN-2004.

PF 15-MAY-2003; 2003US-00440523.

PR 29-DEC-1999; 99US-0173609P.

PR 07-DEC-2000; 2000US-0254414P.

PR 28-DEC-2000; 2000US-00751299.

PR 21-JUN-2001; 2001US-0300189P.

PR 30-JUL-2001; 2001US-0309066P.

PR 22-JAN-2002; 2002US-0351336P.

PR 15-MAY-2002; 2002US-00146772.

PR 09-SEP-2002; 2002US-00241742.

PA (DIVE-) DIVERSA CORP.

PI Desantis G, Short JM, Burk MJ, Wong K, Farwell R, Chatman K;  
 XX WPI; 2004-121569/12.  
 DR N-PSDB; ADI64310.  
 XX Novel isolated or recombinant polypeptide having nitriase activity,  
 PT useful in production of food additives.

XX Claim 46; SEQ ID NO 58; 105bp; English.

CC The invention describes an isolated or recombinant polypeptide (I)  
XX comprising amino acids having a sequence at least 50 % identical to a  
CC sequence (S1) available in electronic form (EC) from the following web  
CC site ftp.segdata.uspto.gov/sequence.html?docid=2004014195, or its  
CC variant, having one or more mutations at residue 55 Lys, Gly or Glu, at  
CC residue 60 glutamic acid, at residue 111 Ser, their combinations or  
CC fragments. (I) is useful for: producing an (R)-ethyl 4-cyano-3-  
CC hydroxybutyric acid; hydrolysing a nitrile to a carboxylic acid;  
CC hydrolysing cyanohydrin moiety or an aminonitrile moiety; producing a  
CC chiral alpha-hydroxy acid molecule or a chiral amino acid molecule;  
CC producing an (S)-ethyl 4-cyano-3-hydroxybutyric acid; producing an (R)-  
CC mandelic acid or an (R)-phenyl lactic acid derivative; modifying a molecule;  
CC derivative or an (R)-phenyl lactic acid derivative; modifying a molecule;  
CC and for identifying a modified compound. The inventive method is useful  
CC for monitoring or determining % enantiomeric excess or % diastereomeric  
CC excess. (I) is useful in the production of food additives and drug  
CC intermediates. This is the amino acid sequence of a nitrilase of the  
CC invention.

XX  
XX Sequence 337 AA;

Query Match 75.9%; Score 1339; DB 8; Length 337;

Best Local Similarity 75.2%; Pred. No. 3.3e-140;

Matches 252; Conservative 30; Mismatches 47; Indels 6; Gaps 1;

QY	1	MKEAIKACVQAAPITVMDLEATVTKTELMEEAARNNARLIAPETWIPGYPMFLMDSP	60
DB	1	VKEAIKACVQAAPITVMDLEATVTKTELMEEAARNNARLIAPETWIPGYPMFLMDSP	60
QY	61	AMAMQFVROYHENSELEIDGPOAKRISDAKRLGIMVTLGMSERVGTLYISOWFIDGND	120
DB	61	AMGQFVRRHENSELVDSPOAKRISDAKRLGIMVTLGMSERVGTLYISOWFIDGND	120
QY	121	TIGARRKLKPTFERTLFGEGDSSSLAVFETSVGRIGLCCWEHLQPLTKALYAONEEI	180
DB	121	TAGARRKLKPTFERTLFGEGDSSSLAVFETSVGRIGLCCWEHLQPLTKALYAONEEI	180
QY	181	HCAAMPSSSLYPNAKALGPDVNVAAASRIYAVEGOCFLASCALVSQSMIDMLCTDEKH	240
DB	181	HFAAMPSSSIYRQATEVIGPEVNVAAASRIYAVEGOCFLASCALVSPEMIMLCTDEKH	240
QY	241	ALLAGGSHSRILIGDGGDLVAPLAENEGILYANLDPGRILAKMAADPAGHYSRPDIT	300
DB	241	SLQAGGYSRIIGDGGDLARPLGENEGILYATLDPARITYAKTADPAGHYSRPDVT	300
QY	301	RLILDRSPKL PVEIEGDLRPYALGKASGTGAOLE	335
DB	301	RLILNRSANQPVVEVGREI-----PASAQGFVE	329

Search completed: April 27, 2006, 00:54:59  
Job time : 133.234 secs

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioceleration Ltd.

OM protein - protein search, using sw model

Run on: April 27, 2006, 01:00:30 ; Search time 33.552 Seconds  
(without alignments)  
830.404 Million cell updates/sec

Title: US-09-751-299-4  
Perfect score: 1765  
Sequence: 1 MKEAIKVCVQAPRIYMDL.....DLRPYLGKASGTGAQLEEI 337

Scoring table: BIOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 572060 seqs, 82675679 residues

Total number of hits satisfying chosen parameters: 572060

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%

Listing first 45 summaries

Database : Issued Patents AA:\*  
1: /cgn2\_6/ptodata/1/1aa/5\_COMB.pep:\*  
2: /cgn2\_6/ptodata/1/1aa/6\_COMB.pep:\*  
3: /cgn2\_6/ptodata/1/1aa/H\_COMB.pep:\*  
4: /cgn2\_6/ptodata/1/1aa/PCTUS\_COMB.pep:\*  
5: /cgn2\_6/ptodata/1/1aa/RE\_COMB.pep:\*  
6: /cgn2\_6/ptodata/1/1aa/backfillset.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Query Length	DB ID	Description
1	955.5	54.1	344	1	US-08-690-493-1 Sequence 1, Appli
2	822	46.6	356	2	US-09-806-876A-2 Sequence 2, Appli
3	776	44.0	369	2	US-09-823-373-5 Sequence 5, Appli
4	776	44.0	369	2	US-09-823-373-14 Sequence 14, Appli
5	638	36.1	354	1	US-08-447-702-5 Sequence 5, Appli
6	638	36.1	354	1	US-08-465-615-5 Sequence 5, Appli
7	152	8.6	267	2	US-09-902-540-14601 Sequence 14601, A
8	139.5	7.9	315	2	US-10-105-294B-2 Sequence 2, Appli
9	139	7.9	311	1	US-07-917-111-5 Sequence 5, Appli
10	139	7.9	311	1	US-07-917-111-6 Sequence 6, Appli
11	139	7.9	311	1	US-08-479-638-5 Sequence 5, Appli
12	139	7.9	311	1	US-08-479-638-6 Sequence 6, Appli
13	139	7.9	311	1	US-08-294-871A-70 Sequence 70, Appli
14	139	7.9	311	2	US-08-876-398A-70 Sequence 70, Appli
15	133	7.5	303	1	US-08-294-871A-46 Sequence 46, Appli
16	133	7.5	303	2	US-08-876-398A-46 Sequence 46, Appli
17	132	7.5	303	1	US-08-294-871A-36 Sequence 36, Appli
18	132	7.5	303	2	US-08-876-398A-36 Sequence 36, Appli
19	131	7.4	303	1	US-08-294-871A-4 Sequence 4, Appli
20	131	7.4	303	1	US-08-294-871A-16 Sequence 16, Appli
21	131	7.4	303	1	US-08-294-871A-34 Sequence 34, Appli
22	131	7.4	303	1	US-08-294-871A-48 Sequence 48, Appli
23	131	7.4	303	1	US-08-294-871A-58 Sequence 58, Appli
24	131	7.4	303	1	US-08-294-871A-68 Sequence 68, Appli
25	131	7.4	303	2	US-08-876-398A-4 Sequence 4, Appli
26	131	7.4	303	2	US-08-876-398A-16 Sequence 16, Appli
27	131	7.4	303	2	US-08-876-398A-34 Sequence 34, Appli

28	131	7.4	303	2	US-08-876-398A-48 Sequence 48, Appli
29	131	7.4	303	2	US-08-876-398A-58 Sequence 58, Appli
30	131	7.4	303	2	US-08-876-398A-68 Sequence 68, Appli
31	130	7.4	303	1	US-08-294-871A-38 Sequence 38, Appli
32	130	7.4	303	1	US-08-294-871A-44 Sequence 44, Appli
33	130	7.4	303	1	US-08-294-871A-50 Sequence 50, Appli
34	130	7.4	303	1	US-08-294-871A-56 Sequence 56, Appli
35	130	7.4	303	1	US-08-294-871A-60 Sequence 60, Appli
36	130	7.4	303	2	US-08-876-398A-38 Sequence 38, Appli
37	130	7.4	303	2	US-08-876-398A-44 Sequence 44, Appli
38	130	7.4	303	2	US-08-876-398A-50 Sequence 50, Appli
39	130	7.4	303	2	US-08-876-398A-56 Sequence 56, Appli
40	130	7.4	303	2	US-08-876-398A-60 Sequence 60, Appli
41	130	7.4	304	1	US-08-900-711-5 Sequence 5, Appli
42	130	7.4	304	1	US-08-815-356-2 Sequence 2, Appli
43	130	7.4	304	1	US-08-415-343B-5 Sequence 5, Appli
44	130	7.4	304	1	US-08-762-433-1 Sequence 1, Appli
45	130	7.4	304	1	US-09-001-219-1 Sequence 1, Appli

ALIGNMENTS

RESULT 1  
US-08-690-493-1  
Sequence 1, Application US/08690493  
Patent No. 5872000  
GENERAL INFORMATION:  
APPLICANT: Yu, Fujio  
TITLE OF INVENTION: No. 5872000el Nitrlase Gene  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Steinberg, Raikin & Davidson, P.C.  
STREET: 1140 Avenue of the Americas  
CITY: New York  
STATE: New York  
COUNTRY: U.S.A.  
ZIP: 10036  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: MS-DOS Editor  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/690,493  
FILING DATE: 31 JUL-1996  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 213061/1995  
FILING DATE: 31-JUL-1995  
ATTORNEY/AGENT INFORMATION:  
NAME: Davidson, Clifford M  
REGISTRATION NUMBER: 32,728  
REFERENCE/DOCKET NUMBER: 3821005  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (212) 768-3800  
TELEFAX: (212) 382-2124  
INFORMATION FOR SEQ ID NO: 1:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 344 amino acid residues  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
ORIGINAL SOURCE:  
ORGANISM: Gordona terrae  
STRAIN: MA-1  
CELL TYPE: unicellular organism  
FEATURE:  
OTHER INFORMATION: Xaa is Met or a deletion  
US-08-690-493-1  
Query Match 54.1%; Score 955.5; DB 1; Length 344;  
Best Local Similarity 58.3%; Pred. No. 2,4e-97;  
Matches 182; Conservative 47; Mismatches 82; Indels 1;



;; CURRENT APPLICATION NUMBER: US/09/823,373  
;; CURRENT FILING DATE: 2001-03-30  
;; PRIOR APPLICATION NUMBER: 60/193,707  
;; PRIOR FILING DATE: 2000-03-31  
;; NUMBER OF SEQ ID NOS: 32  
;; SOFTWARE: Microsoft Office 97  
;; SEQ ID NO 14  
;; LENGTH: 369  
;; TYPE: PRT  
;; ORGANISM: Acidovorax facilis  
US-09-823-373-14

Query Match 44.0%; Score 776; DB 2; Length 369;  
Best Local Similarity 47.6%; Pred. No. 2, 4e-77;

Matches 150; Conservative 61; Mismatches 100; Indels 4; Gaps 3;

QY 8 ACVQAPRYMDLEATVDKTIELMEBAARNRNLIAPEETWPGYPMFLDSDPAMQPV 67  
DB 11 ATVQAPRYMDLEATVDKTIELMEBAARNRNLIAPEETWPGYPMFLDSDPAMQPV 70  
QY 68 ROYHNSLELDGQAKRISDAKRLGIMVTLGMSERVGTLYSQWFIQDNGDTIGARRK 127  
DB 71 SRHNSLELDGQAKRISDAKRLGIMVTLGMSERVGTLYSQWFIQDNGDTIGARRK 130  
QY 128 LKPTFVERTLFGEGSSLAVFETSVRLGGLCCMHLOPLTKYALYAONEBHCAMPS 187  
DB 131 LKPTFVERTLFGEGSSLAVFETSVRLGGLCCMHLOPLTKYALYAONEBHCAMPS 190  
QY 188 FG-LYNAKALGPDVNVASRIYAVEGCFVLASCALVSQSMIDMLCTDDEKHALLLAG 246  
DB 191 MSPLQPDVFQ-LSEIANATVTRSYALEGQTFVLCSTQVIGPSAIEFCLNDEQRALLPQG 249  
QY 247 GGHRSRIGDGDVLAFLAENEGILYANLDPGVRLAKMAADPAHYSRPDTLLILR 306  
DB 250 CCMARYGPDGSELAKPLADEAGILYASIDLEQIILLARAGADPVGHYSRPDVLVSQFDP 309  
QY 307 SPKL.PV--VEIEGDL 319  
DB 310 RHHTPVHRIGIDRL 324

RESULT 5  
US-08-447-702-5  
;; Sequence 5, Application US/08447702  
;; Patent No. 5629190  
;; GENERAL INFORMATION:  
;; APPLICANT: Petre, Dominique  
;; APPLICANT: Cerbeled, Edith  
;; APPLICANT: Levy-Schill, Sophie  
;; APPLICANT: Crouzet, Joel  
;; TITLE OF INVENTION: Polypeptides Possessing A Nitriase  
;; TITLE OF INVENTION: Activity, DNA Sequence Coding For Said Polypeptides,  
Patent No. 5629190  
;; TITLE OF INVENTION: Expression Cassettes and Host Microorganisms Enabling Them  
;; TITLE OF INVENTION: to be Obtained, and Method of Converting Nitrites to Carboxy-  
TITLE OF INVENTION: Lates by Means of Said Polypeptides  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: BURNS, DOANE, SWECKER & MATHIS  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: VA  
COUNTRY: USA  
ZIP: 22313-1404  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/447,702  
FILING DATE: 23-MAY-1995  
CLASSIFICATION: 530

;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: US 08/194,588  
;; FILING DATE: 10-FEB-1994  
;; PRIOR APPLICATION DATA:  
;; APPLICATION NUMBER: FR 92-09-882  
;; FILING DATE: 10-AUG-1992  
;; ATTORNEY/AGENT INFORMATION:  
NAME: McGowan, Malcolm K.  
REGISTRATION NUMBER: 39,300  
REFERENCE/DOCKET NUMBER: 003025-019  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 703-836-2021  
TELEFAX: 703-836-6620  
;; INFORMATION FOR SEQ ID NO: 5:  
;; SEQUENCE CHARACTERISTICS:  
;; LENGTH: 354 amino acids  
;; TYPE: amino acid  
;; TOPOLOGY: linear  
;; MOLECULE TYPE: protein  
US-08-447-702-5

Query Match 36.1%; Score 638; DB 1; Length 354;  
Best Local Similarity 41.2%; Pred. No. 5e-62;  
Matches 134; Conservative 62; Mismatches 107; Indels 22; Gaps 5;

QY 5 IKVACVQAPRYMDLEATVDKTIELMEBAARNRNLIAPEETWPGYPMFLDSDPAMQPV 59  
DB 7 VVQVQAPRYMDLEATVDKTIELMEBAARNRNLIAPEETWPGYPMFLDSDPAMQPV 66  
QY 60 PAMQPFVRYHNSLELDGQAKRISDAKRLGIMVTLGMSERVGTLYSQWFIQDNG 119  
DB 67 MMVAFLF-----KNIEIPSKVEVQISDAKKNVYVCVSSEKONASLYLQWLPDNG 121  
QY 120 DTIGARRKLPFVERTLFGEGSSLAVFETSVRLGGLCCMHLOPLTKYALYAONEB 179  
DB 122 NLIGHRRKFKPTSSRAVWGDDGSMARVFKTEYGNLGLQCMEHALPLINIAMGSLNQ 181  
QY 180 IHCAMPSPSLYPNAXKALGPDVNVAA-----SRIVYEGGCFVLASCALVSQSMI 230  
DB 182 VHVASMPAP--VPKAVSSRVSSVCASTNANHQIISQFYAISNVYVIMSTNLVGQDMT 239  
QY 231 DMLCTDDEKHALLLAGGHSRITGPDGDVLAFLAENEGILYANLDPGVRLAKMAAD 290  
DB 240 DMIGDESKNPLPGSGNTALIS-NTGEIILSIQDAEGLAVAILDNLQIYGMKLD 298  
QY 291 AGHYSRPDTLLILDRSPKL.PVVEI 315  
DB 299 AGHYSTPGRLSLTFDQSEHVPVKI 323

RESULT 6  
US-08-465-615-5  
;; Sequence 5, Application US/08465615  
;; Patent No. 5635391  
;; GENERAL INFORMATION:  
;; APPLICANT: Petre, Dominique  
;; APPLICANT: CERBELED, Edith  
;; APPLICANT: LEVY-SCHILL, Sophie  
;; APPLICANT: CROUZET, Joel  
;; TITLE OF INVENTION: POLYPEPTIDES POSSESSING A NITRIASE  
;; TITLE OF INVENTION: ACTIVITY, DNA SEQUENCE CODING FOR SAID POLYPEPTIDES,  
TITLE OF INVENTION: TO BE OBTAINED, AND METHOD OF CONVERTING NITRIDES TO  
TITLE OF INVENTION: CARBOXYLATES BY MEANS OF SAID POLYPEPTIDE  
NUMBER OF SEQUENCES: 5  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Burns, Doane, Swecker & Mathis  
STREET: P.O. Box 1404  
CITY: Alexandria  
STATE: Virginia  
COUNTRY: United States  
ZIP: 22313-1404  
COMPUTER READABLE FORM:



Db 134 ECLPNOHLEKRYREGD-LGFGVDFHGVQVGMCLCNDRRMDEVYRSIALQGAELVIVGX 192  
QY 186 -----PFSLYPNNAKALGPDVNVAASTIYAVEGOCFLASCALVSQSMIDMLCTDDEK 239  
Db 193 NTPDFVPMQOEPPHAKMF-----THLISLOAGAYQNSVFVAAL-----GKSGF-----EBOH 239  
QY 240 HALLAGGSHRIIGPDGDVLVAPLAENEGILY-ANLDPGVRIILAKMAADPAGHYSPD 298  
Db 240 HMT-----GGSAAAPSGEITLAKAAGEDEVVVKRADIWG-KPYKESVFDPAAH-RRPD 292  
QY 299 ITRLLIDRSPPK---LPV 312  
Db 293 AVGIABERKRGAPLPV 309

RESULT 9  
US-07-917-111-5  
; Sequence 5, Application US/07917111  
; Patent No. 5565344  
; GENERAL INFORMATION:  
; APPLICANT: Nanda, Hirokazu  
; APPLICANT: Yamada, Yukio  
; APPLICANT: Takano, Masayuki  
; APPLICANT: Ikenaka, Yasuhiro  
; APPLICANT: Takahashi, Satomi  
; APPLICANT: Yajima, Kazuyoshi  
; TITLE OF INVENTION: PROCESS FOR PRODUCTION OF D-ALPHA-AMINO  
; TITLE OF INVENTION: ACIDS  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Wegner, Cantor, Mueller & Player  
; STREET: 1233 20th Street, N.W., Suite 300  
; CITY: Washington  
; STATE: D.C.  
; ZIP: 20036-8218  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/917,111  
; FILING DATE: 19920807  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 400848/1990  
; FILING DATE: 07-DEC-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 407922/1990  
; FILING DATE: 27-DEC-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 078840/1991  
; FILING DATE: 11-APR-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/JP91/01696  
; FILING DATE: 06-DEC-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Player Esq., William E.  
; REGISTRATION NUMBER: 31,409  
; REFERENCE/DOCKET NUMBER: P-500-23486  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: 202-887-0400  
; TELEFAX: 202-835-0605  
; TELERX: 440706  
; INFORMATION FOR SEQ ID NO: 5:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 311 amino acids  
; TYPE: AMINO ACID  
; TOPOLOGY: linear  
; MOLECULE TYPE: protein  
; US-07-917-111-5

Query Match 7.9%; Score 139; DB 1; Length 311;

Best Local Similarity 23.7%; Pred. No. 1,2e-06;  
Matches 80; Conservative 49; Mismatches 146; Indels 62; Gaps 15;

QY 5 IKVACVQAPAIYMD-LEATYDKTIELMEBAARNRARIAPETWIPGYPMFLMDSAW 62  
Db 4 VNAAAQCMGPIRSRETRKDTVRRLIALMREAKAGSDLVVFTEALTLTF-----FPRW 56  
QY 63 AMQFVQVHNSLSD-----GPOAKRISDAKRLGIWVIGMSF--RVGG--TLXI 110  
Db 57 VI-----EDEALDSFEYKEMPGETOPLFDEAKRLIEFYIGYALAEBOGKRFPN 109  
QY 111 SQWFIGDNGDTIGARRKL-----KPTVEVTLFGEKGGSSIAVEFETSGRGLG 158  
Db 110 TSLVDSGRIVGKYRKVHLPGHKEPDGRGHGLEKRYFEFGD-LGFGVARAFDGVWGM 168  
QY 159 LCCWEHLQPLTKVLYAONEIHCAM--PFSLYPNNAKALGPDVNVAASTIYAVEGOC 216  
Db 169 CINDRRMPEFYRVWGLQGVEMWMLGYWTPYDHTGHDDIDSLTQFHNHLSMQAGAYONST 228  
QY 217 FVLASCALVSQSMIDMLCTDDEKHALLAGGSHRIIGPDGDVLVAPLAENEGILY-- 274  
Db 229 VWIGTAK-----CGTEGSKMV---GQSVIYAP-SGRIYMAACTIEDTLTARC 273  
QY 275 NLDPGVRIILAKMAADPAGHYSPDITRLLIDRSPLP 311  
Db 274 DLDWGR-YRETIFDPARH-REPDAYRLIVRKGAVP 308

RESULT 10  
US-07-917-111-6  
; Sequence 6, Application US/07917111  
; Patent No. 5565344  
; GENERAL INFORMATION:  
; APPLICANT: Nanda, Hirokazu  
; APPLICANT: Yamada, Yukio  
; APPLICANT: Takano, Masayuki  
; APPLICANT: Ikenaka, Yasuhiro  
; APPLICANT: Takahashi, Satomi  
; APPLICANT: Yajima, Kazuyoshi  
; TITLE OF INVENTION: PROCESS FOR PRODUCTION OF D-ALPHA-AMINO  
; TITLE OF INVENTION: ACIDS  
; NUMBER OF SEQUENCES: 6  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Wegner, Cantor, Mueller & Player  
; STREET: 1233 20th Street, N.W., Suite 300  
; CITY: Washington  
; STATE: D.C.  
; ZIP: 20036-8218  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/07/917,111  
; FILING DATE: 19920807  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 400848/1990  
; FILING DATE: 07-DEC-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 407922/1990  
; FILING DATE: 27-DEC-1990  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: JP 078840/1991  
; FILING DATE: 11-APR-1991  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: PCT/JP91/01696  
; FILING DATE: 06-DEC-1991  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Player Esq., William E.  
; REGISTRATION NUMBER: 31,409  
; REFERENCE/DOCKET NUMBER: P-500-23486





NUMBER OF SEQUENCES: 6  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Wegner, Cantor, Mueller & Player  
STREET: 1233 20th Street, N.W., Suite 300  
CITY: Washington  
STATE: D.C.  
ZIP: 20036-8218  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/479,638  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/917,111  
FILING DATE: 07-AUG-1992  
APPLICATION NUMBER: JP 400848/1990  
FILING DATE: 07-DEC-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 407922/1990  
FILING DATE: 27-DEC-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 078840/1991  
FILING DATE: 11-APR-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/JP91/01696  
FILING DATE: 06-DEC-1991  
ATTORNEY/AGENT INFORMATION:  
NAME: Player Esq., William E.  
REGISTRATION NUMBER: 31,409  
REFERENCE/DOCKET NUMBER: P-500-23486  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: 202-887-0400  
TELEFAX: 202-835-0605  
TELEX: 440706  
INFORMATION FOR SEQ ID NO: 6:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 311 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
ORGANISM: Pseudomonas  
STRAIN: KNK 003A (FERM BP-3181)  
US-08-479-638-6

Query Match 7.9%; Score 139; DB 1; Length 311;  
Best Local Similarity 23.7%; Pred. No. 1.2e-06;  
Matches 80; Conservative 49; Mismatches 146; Indels 62; Gaps 15;

5 IKVACVQAPRYMD--LEATVDKTEIMBEAARNNAKRLAEPBTWIPGYPFLMLNDSPW 62  
4 VNAAAQWQPIRSRSETRKQIVRRLLALMBEARRGSDLVFTELATTE-----FPRW 56  
63 AMQFVQYVENSLELD-----GPOAKRISDAKRLGIWTLGMSR--RVGG--TLYI 110  
57 VI-----EDAEHLDSFYEKEMPGEBTQPLDPEAKRLTIGFLYGAELAEBSGRARRN 109  
111 SCWFIGDNGDTTGARRKL-----KPTFVERTLFGESGSSSLAVFETSVGRLLG 158  
110 TSLIVDRSGRIYGVKRYKVLPGHKEPQGRKQHLEKRYFEPD-LGFGVWRAFDGVCM 168  
159 LCCMEHLQVLTAKALYAQNEEIHCAW--PSSLYNNAKALGPDVNVAAASRIYAEQGC 216  
169 CICNDRWPEYRVMGLQGVEMVGLGYNTPYDHTGDDIDSLTQFINHLSMQAGAYONST 228  
217 FYLASCALVSQSMIDMLCTDDEKHALLAGGSHRIIGPDGGLVAPLAENEGILYA-- 274  
229 WVIQTAK-----CGTEBSKRV-----GQSVIYAP-SGEIVAMACTIDEIITARC 273

275 NLDPGVRIAKMAADPAGHSRPOITRLILDRSPKL 311  
274 DLDWGR-YRETIYFPAH-REPDAYRIVERKGV 308

US-08-294-871A-70  
Sequence 70, Application US/08294871A  
Patent No. 5824522  
GENERAL INFORMATION:  
APPLICANT: Ikenaka, Yasuhiro  
APPLICANT: Namba, Hirokazu  
APPLICANT: Takano, Masayuki  
APPLICANT: Yajima, Kazuyoshi  
APPLICANT: Yamada, Yukio  
APPLICANT: Takahashi, Satomi  
APPLICANT: Okubo, Kazuma  
APPLICANT: Yamada, Kazuhiro  
TITLE OF INVENTION: Immobilized Enzyme Preparation and  
TITLE OF INVENTION: Process for Producing D-a-Amino Acid  
NUMBER OF SEQUENCES: 70  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Foley & Lardner  
STREET: 3000 K Street, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patentin Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/294,871A  
FILING DATE: 22-AUG-1994  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/971,758  
FILING DATE: 12-APR-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 07/917,111  
FILING DATE: 07-AUG-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/211,641  
FILING DATE: 11-APR-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 400848/1990  
FILING DATE: 07-DEC-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/JP91/01696  
FILING DATE: 06-DEC-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 407922/1990  
FILING DATE: 27-DEC-1990  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 078840/1991  
FILING DATE: 11-APR-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 140051/1991  
FILING DATE: 12-JUN-1991  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/JP92/00739  
FILING DATE: 10-JUN-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 212692/1992  
FILING DATE: 10-AUG-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: PCT/JP93/01101  
FILING DATE: 05-AUG-1993  
PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP 340078/1992  
FILING DATE: 21-DEC-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Wegner, Harold C.  
REGISTRATION NUMBER: 25,258  
REFERENCE/DOCKET NUMBER: 74129/127/AOPA  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 672-5300  
TELEFAX: (202) 672-5399  
TELEX: 904136  
INFORMATION FOR SEQ ID NO: 70:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 311 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-294-871A-70

Query Match 7.9%; Score 139; DB 1; Length 311;  
Best Local Similarity 23.7%; Pred. No. 1.2e-06;  
Matches 80; Conservative 49; Mismatches 146; Indels 62; Gaps 15;

QY 5 IKVACVQAPAYMD--LEATVDKTIELMEBAARNNAFLAEPETWIPGYPMFLWLDSPAW 62  
DB 4 VNAAAQMGPIRSRSETRKDTYRLIALMREKARGSDLVVFTLALTF-----FPRW 56  
QY 63 AMQFVROYHNSLELD-----GPOAKRISDAKRLGIMVTLGMSR--RVGG--TLVI 110  
DB 57 VI-----EDBALDSFYEKEMPGPTOPLFDEAKKLEIGFYGLVYALAEAGGKKRFN 109  
QY 111 SQMFIGNDGTIGARRKL-----KPTFVERTLFGEGDSSLAVFETSVGRIGG 158  
DB 110 TSLIVDRSGRIYGVKRYKHLPGHKEPQGRKHQHLKRYFEPGD--LGFGVWRAPDGVWGM 168  
QY 159 LCCWEHLQPLTKYALYAQNEEIHCAAM--PFSIYPNAKALGPDVNVAAASRIYAVEGQC 216  
DB 169 CICNDRWPETRYVMGLQGVEMVMGLGYNTPYDHTGHDIDSLTQFNHLSWQAGAYONST 228  
QY 217 FYLASCALVSQSMIDMLCTDDEKHALLLAGGHSRIIGPDGDLVAPLAENEGLIYA-- 274  
DB 229 WITGTAK-----CGTEEGSKMV---GQSYIVAP--SGEIVYVMACTIEDIITARC 273  
QY 275 NIDPGVRIIAKMAADPAGHYSRPDTIRLLIDRSPKLP 311  
DB 274 DLDMGKR--YRETIFDPARH--REPDAYRLIVERKGAVP 308

RESULT 14  
US-08-876-398A-70  
Sequence 70, Application US/08876398A  
Patent No. 6083752  
GENERAL INFORMATION:  
APPLICANT: IKENAKA, Yasuhiro  
APPLICANT: NAKANO, Hirokazu  
APPLICANT: YAJIMA, Kazuyoshi  
APPLICANT: YAMADA, Yukio  
APPLICANT: TAKAHASHI, Satomi  
TITLE OF INVENTION: DNA CODING FOR DECARBAMYLASE IMPROVED IN  
NUMBER OF SEQUENCES: 70  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: FOLEY & LARDNER  
STREET: 3000 K Street, N.W.  
CITY: Washington  
STATE: D.C.  
COUNTRY: U.S.A.  
ZIP: 20007-5109  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30

CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/876,398A  
FILING DATE: 16-JUN-1997  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/211,641  
FILING DATE: 11-APR-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: WO PCT/JP93/01101  
FILING DATE: 05-AUG-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 340078/1992  
FILING DATE: 21-DEC-1992  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: JP 212692/1992  
FILING DATE: 10-AUG-1992  
ATTORNEY/AGENT INFORMATION:  
NAME: Wegner, Harold C.  
REGISTRATION NUMBER: 25,258  
REFERENCE/DOCKET NUMBER: 74129/130  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (202) 672-5300  
TELEFAX: (202) 672-5399  
INFORMATION FOR SEQ ID NO: 70:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 311 amino acids  
TYPE: amino acid  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-876-398A-70

Query Match 7.9%; Score 139; DB 2; Length 311;  
Best Local Similarity 23.7%; Pred. No. 1.2e-06;  
Matches 80; Conservative 49; Mismatches 146; Indels 62; Gaps 15;

QY 5 IKVACVQAPAYMD--LEATVDKTIELMEBAARNNAFLAEPETWIPGYPMFLWLDSPAW 62  
DB 4 VNAAAQMGPIRSRSETRKDTYRLIALMREKARGSDLVVFTLALTF-----FPRW 56  
QY 63 AMQFVROYHNSLELD-----GPOAKRISDAKRLGIMVTLGMSR--RVGG--TLVI 110  
DB 57 VI-----EDBALDSFYEKEMPGPTOPLFDEAKKLEIGFYGLVYALAEAGGKKRFN 109  
QY 111 SQMFIGNDGTIGARRKL-----KPTFVERTLFGEGDSSLAVFETSVGRIGG 158  
DB 110 TSLIVDRSGRIYGVKRYKHLPGHKEPQGRKHQHLKRYFEPGD--LGFGVWRAPDGVWGM 168  
QY 159 LCCWEHLQPLTKYALYAQNEEIHCAAM--PFSIYPNAKALGPDVNVAAASRIYAVEGQC 216  
DB 169 CICNDRWPETRYVMGLQGVEMVMGLGYNTPYDHTGHDIDSLTQFNHLSWQAGAYONST 228  
QY 217 FYLASCALVSQSMIDMLCTDDEKHALLLAGGHSRIIGPDGDLVAPLAENEGLIYA-- 274  
DB 229 WITGTAK-----CGTEEGSKMV---GQSYIVAP--SGEIVYVMACTIEDIITARC 273  
QY 275 NIDPGVRIIAKMAADPAGHYSRPDTIRLLIDRSPKLP 311  
DB 274 DLDMGKR--YRETIFDPARH--REPDAYRLIVERKGAVP 308

RESULT 15  
US-08-294-871A-46  
Sequence 46, Application US/08294871A  
Patent No. 5824522  
GENERAL INFORMATION:  
APPLICANT: IKENAKA, Yasuhiro  
APPLICANT: NAKANO, Hirokazu  
APPLICANT: YAJIMA, Kazuyoshi  
APPLICANT: YAMADA, Yukio  
APPLICANT: TAKAHASHI, Satomi  
APPLICANT: OKUBO, Kazuhiko



**THIS PAGE BLANK (USPTO)**

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: April 27, 2006, 01:01:20 ; Search time 109.044 Seconds  
(without alignments)  
1291.301 Million cell updates/sec

Title: US-09-751-299-4

Perfect score:

Sequence: 1 MKEAIKVACVQAPIYMDL...DLRPYALGKASETGAQLEEI 337

### Scoring table:

Gapop 10.0 , Gapext 0.5

Searched: 1867569 seqs, 417829326 residues

Total number of hits satisfying chosen parameters: 1867569

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Database : Published\_Applications\_AA\_Main:\*

- |    |   |
|----|---|
| 1: | /cgn2_6/ptodatta/1/pmbpaa/US07_PUBCOMB.rep.*  |
| 2: | /cgn2_6/ptodatta/1/pmbpaa/US08_PUBCOMB.rep.*  |
| 3: | /cgn2_6/ptodatta/1/pmbpaa/US09_PUBCOMB.rep.*  |
| 4: | /cgn2_6/ptodatta/1/pmbpaa/US10A_PUBCOMB.rep.* |
| 5: | /cgn2_6/ptodatta/1/pmbpaa/US10B_PUBCOMB.rep.* |
| 6: | /cgn2_6/ptodatta/1/pmbpaa/US11_PUBCOMB.rep.*  |

**Pred. No.** is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed and is derived by analysis of the total score distribution.

## SUMMARIES

Result	Score	Query	Match	Length	DB	ID	Description
No.							
1	1765	100.0	337	3	US-09-751-299-4	Sequence 4, Appl1	
2	1761	99.8	337	4	US-10-146-772-386	Sequence 386, Appl	
3	1761	99.8	337	4	US-10-241-742-386	Sequence 386, Appl	
4	1761	99.8	337	4	US-10-440-523-386	Sequence 386, Appl	
5	1761	99.8	337	4	US-10-440-503-386	Sequence 386, Appl	
6	1761	99.8	337	4	US-10-461-925-386	Sequence 386, Appl	
7	1339	75.9	337	4	US-10-146-772-48	Sequence 48, Appl	
8	1339	75.9	337	4	US-10-146-772-58	Sequence 58, Appl	
9	1339	75.9	337	4	US-10-241-742-48	Sequence 48, Appl	
10	1339	75.9	337	4	US-10-241-742-58	Sequence 58, Appl	
11	1339	75.9	337	4	US-10-440-523-48	Sequence 48, Appl	
12	1339	75.9	337	4	US-10-440-523-58	Sequence 58, Appl	
13	1339	75.9	337	4	US-10-440-503-48	Sequence 48, Appl	
14	1339	75.9	337	4	US-10-440-503-58	Sequence 58, Appl	
15	1339	75.9	337	4	US-10-461-925-48	Sequence 48, Appl	
16	1339	75.9	337	4	US-10-461-925-58	Sequence 58, Appl	
17	1339	74.5	333	4	US-10-146-772-292	Sequence 292, Appl	
18	1139	64.5	333	4	US-10-241-742-292	Sequence 292, Appl	
19	1139	64.5	333	4	US-10-440-523-292	Sequence 292, Appl	
20	1139	64.5	333	4	US-10-440-503-292	Sequence 292, Appl	
21	1139	64.5	333	4	US-10-461-925-292	Sequence 292, Appl	
22	1136	64.4	332	4	US-10-146-772-140	Sequence 140, Appl	
23	1136	64.4	332	4	US-10-241-742-140	Sequence 140, Appl	
24	1136	64.4	332	4	US-10-440-523-140	Sequence 140, Appl	
25	1136	64.4	332	4	US-10-440-503-140	Sequence 140, Appl	
26	1136	64.4	332	4	US-10-461-925-140	Sequence 140, Appl	
27	1091	61.8	355	4	US-10-146-772-306	Sequence 306, Appl	

28	1091	61.8	355	4	US-10-241-742-306	Sequence 306, App
29	1091	61.8	355	4	US-10-440-523-306	Sequence 306, App
30	1091	61.8	355	4	US-10-440-503-106	Sequence 306, App
31	1091	61.8	355	4	US-10-461-925-106	Sequence 306, App
32	1081	61.2	345	4	US-10-148-772-334	Sequence 334, App
33	1081	61.2	345	4	US-10-241-742-334	Sequence 334, App
34	1081	61.2	345	4	US-10-440-523-334	Sequence 334, App
35	1081	61.2	345	4	US-10-440-503-334	Sequence 334, App
36	1081	61.2	345	4	US-10-461-925-334	Sequence 334, App
37	1051	59.5	334	4	US-10-146-742-6	Sequence 6, App1
38	1051	59.5	334	4	US-10-241-742-6	Sequence 6, App1
39	1051	59.5	334	4	US-10-440-523-6	Sequence 6, App1
40	1051	59.5	334	4	US-10-440-503-6	Sequence 6, App1
41	1051	59.5	334	4	US-10-461-925-6	Sequence 6, App1
42	1042	59.0	353	4	US-10-146-772-212	Sequence 212, App
43	1042	59.0	353	4	US-10-241-742-212	Sequence 212, App
44	1042	59.0	353	4	US-10-440-523-212	Sequence 212, App
45	1042	59.0	353	4	US-10-440-503-212	Sequence 212, App

## ALIGNMENTS

	RESULT 1	
	US-09-751-299-4	
	; Sequence 4, Application US/09751299	
	; Patent No. US20020012974A1	
	GENERAL INFORMATION:	
	APPLICANT: Madden, Mark	
	APPLICANT: Weiner, David P.	
	APPLICANT: Chaplin, Jennifer A.	
	TITLE OF INVENTION: METHODS FOR PRODUCING ENANTIOMERICALLY PURE	
	TITLE OF INVENTION: ALPHA-SUBSTITUTED CARBOXYLIC ACIDS	
	FILE REFERENCE: DIVERI440-2	
	CURRENT APPLICATION NUMBER: US/09/751,299	
	CURRENT FILING DATE: 2000-12-29	
	PRIOR APPLICATION NUMBER: 60/254,414	
	PRIOR FILING DATE: 2000-12-07	
	PRIOR APPLICATION NUMBER: 60/173,609	
	PRIOR FILING DATE: 1999-12-29	
	NUMBER OF SEQ ID NOS: 4	
	SOFTWARE: PatentIn Ver. 2.1	
	SEQ ID NO 4	
	LENGTH: 337	
	TYPE: PRF	
	ORGANISM: Unknown Organism	
	FEATURES:	
	OTHER INFORMATION: Description of Unknown Organism: Obtained from an	
	OTHER INFORMATION: environmental sample	
	US-09-751-299-4	
QY	Query Match	100.0%; Score 1765; DB 3; Length 337;
	Best Local Similarity	100.0%; Pred. No. 1.3e-176;
	Matches 337; Conservative	0; Mismatches 0; Indels 0; Gaps 0
Dd	1 MKSAIVACVQAIPYMDEATVDVTIELMBSAARNRRLTAFPEWIPGYPMTLMDSP	60
	1 MKSAIVACVQAIPYMDEATVDVTIELMBSAARNRRLTAFPEWIPGYPMTLMDSP	60
QY	61 AMAQEFVROYHENSLELDPOAKRISDAKRLIGIMVTLGMSERVGGLTYSOWFIGDNGD	120
	61 AMAQEFVROYHENSLELDPOAKRISDAKRLIGIMVTLGMSERVGGTLTYSOWFIGDNGD	120
Dd	121 TIGARRKLKPTPERLTLPBEGSGSSLAIVPFSVGRIGLCCEHHLOPLTKALYAQNBEI	180
	121 TIGARRKLKPTPERLTLPBEGSGSSLAIVPFSVGRIGLCCEHHLOPLTKALYAQNBEI	180
QY	121 TIARRRKLKPTPERLTLPBEGSGSSLAIVPFSVGRIGLCCEHHLOPLTKALYAQNBEI	180
	121 TIARRRKLKPTPERLTLPBEGSGSSLAIVPFSVGRIGLCCEHHLOPLTKALYAQNBEI	180
Dd	181 HCAAMPFSLSLPPAAKALGPDVNVVAASRIYAVEGCFFLASCALVSOSMTIDMLCTDDEKH	240
	181 HCAAMPFSLSLPPAAKALGPDVNVVAASRIYAVEGCFFLASCALVSOSMTIDMLCTDDEKH	240
QY	241 ALLAGGSHRIITGPDGDVLVAPLAENEGILYANDPGVRILAKMAADPAHYSRPDIT	300
	241 ALLAGGSHRIITGPDGDVLVAPLAENEGILYANDPGVRILAKMAADPAHYSRPDIT	300

Db 241 ALLIAGGHSRIIGPDGDLVAPLAENEGILYANLDPGVRIILAKMAADPAHYSRPDIT 300  
Qy 301 RLIDRSPKLPVEIEGDLRPYALGKASETGAOLEEI 337  
Db 301 RLIDRSPKLPVEIEGDLRPYALGKASETGAOLEEI 337

## RESULT 2

US-10-146-772-386  
; Sequence 386, Application US/10146772  
; Publication No. US20030124698A1  
; GENERAL INFORMATION:  
; APPLICANT: Short, Jay  
; APPLICANT: Weiner, David  
; APPLICANT: Chaplin, Jennifer  
; APPLICANT: Chi, Ellen  
; APPLICANT: Milan, Aileen  
; APPLICANT: Desantis, Grace  
; APPLICANT: Madden, Mark  
; APPLICANT: Burt, Mark  
; TITLE OF INVENTION: Nitri1ases  
; FILE REFERENCE: Docket No. US20030124698A1 DIV-013US  
; CURRENT APPLICATION NUMBER: US/10/146,772  
; CURRENT FILING DATE: 2002-05-15  
; PRIOR APPLICATION NUMBER: US 60/309,006  
; PRIOR FILING DATE: 2001-07-30  
; PRIOR APPLICATION NUMBER: US 60/351,336  
; PRIOR FILING DATE: 2002-01-22  
; PRIOR APPLICATION NUMBER: US 60/300,189  
; PRIOR FILING DATE: 2001-06-21  
; PRIOR APPLICATION NUMBER: US 09/751,299  
; PRIOR FILING DATE: 2000-12-28  
; PRIOR APPLICATION NUMBER: US 60/254,414  
; PRIOR FILING DATE: 2000-12-07  
; PRIOR APPLICATION NUMBER: US 60/173,609  
; PRIOR FILING DATE: 1999-12-29  
; NUMBER OF SEQ ID NOS: 386  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 386  
; LENGTH: 337  
; TYPE: PRT  
; ORGANISM: Unknown  
; FEATURE:  
; OTHER INFORMATION: Obtained from an environmental sample  
US-10-146-772-386

Query Match 99.8%; Score 1761; DB 4; Length 337;  
Best Local Similarity 99.7%; Pred. No. 3.5e-176;  
Matches 336; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MKEAIKVAQVQADPIYMDLEATVDKTELMEEAARNNARLIAPETWIPGYPMFLMDSP 60  
Db 1 MKEAIKVAQVQADPIYMDLKATVDKTELMEEAARNNARLIAPETWIPGYPMFLMDSP 60  
Qy 61 AMAQPFQYQYHENSLELDGPOAKRISDAKRLGIMVTLGMSERVGTLTYSQWFIQNGD 120  
Db 61 AMAQPFQYQYHENSLELDGPOAKRISDAKRLGIMVTLGMSERVGTLTYSQWFIQNGD 120  
Qy 121 TIGARRKLKPTFVERTLFGEGDSSSLAVFETSVGRIGLCCWEHLQPLTKYALYAONEEI 180  
Db 121 TIGARRKLKPTFVERTLFGEGDSSSLAVFETSVGRIGLCCWEHLQPLTKYALYAONEEI 180  
Qy 181 HCAAMPFSIYPNAKALGPDVNVAAASRIYAVEGQCFVLASCALVSQSMIDMLCTDDEKH 240  
Db 181 HCAAMPFSIYPNAKALGPDVNVAAASRIYAVEGQCFVLASCALVSQSMIDMLCTDDEKH 240  
Qy 241 ALLIAGGHSRIIGPDGDLVAPLAENEGILYANLDPGVRIILAKMAADPAHYSRPDIT 300  
Db 241 ALLIAGGHSRIIGPDGDLVAPLAENEGILYANLDPGVRIILAKMAADPAHYSRPDIT 300  
Qy 301 RLIDRSPKLPVEIEGDLRPYALGKASETGAOLEEI 337  
Db 301 RLIDRSPKLPVEIEGDLRPYALGKASETGAOLEEI 337

RESULT 3  
US-10-241-742-386  
; Sequence 386, Application US/10241742  
; Publication No. US20040002147A1  
; GENERAL INFORMATION:  
; APPLICANT: Short, Jay  
; APPLICANT: Weiner, David  
; APPLICANT: Chaplin, Jennifer  
; APPLICANT: Chi, Ellen  
; APPLICANT: Milan, Aileen  
; APPLICANT: Desantis, Grace  
; APPLICANT: Madden, Mark  
; APPLICANT: Burt, Mark  
; TITLE OF INVENTION: Nitri1ases  
; FILE REFERENCE: Docket No. US20040002147A1 DIV-013US  
; CURRENT APPLICATION NUMBER: US/10/241,742  
; CURRENT FILING DATE: 2002-09-09  
; PRIOR APPLICATION NUMBER: US/10/146,772  
; PRIOR FILING DATE: 2002-05-15  
; PRIOR APPLICATION NUMBER: US 60/309,006  
; PRIOR FILING DATE: 2001-07-30  
; PRIOR APPLICATION NUMBER: US 60/351,336  
; PRIOR FILING DATE: 2002-01-22  
; PRIOR APPLICATION NUMBER: US 60/300,189  
; PRIOR FILING DATE: 2001-06-21  
; PRIOR APPLICATION NUMBER: US 09/751,299  
; PRIOR FILING DATE: 2000-12-28  
; PRIOR APPLICATION NUMBER: US 60/254,414  
; PRIOR FILING DATE: 2000-12-07  
; PRIOR APPLICATION NUMBER: US 60/173,609  
; PRIOR FILING DATE: 1999-12-29  
; NUMBER OF SEQ ID NOS: 386  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 386  
; LENGTH: 337  
; TYPE: PRT  
; ORGANISM: Unknown  
; FEATURE:  
; OTHER INFORMATION: Obtained from an environmental sample  
US-10-241-742-386

Query Match 99.8%; Score 1761; DB 4; Length 337;  
Best Local Similarity 99.7%; Pred. No. 3.5e-176;  
Matches 336; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MKEAIKVAQVQADPIYMDLEATVDKTELMEEAARNNARLIAPETWIPGYPMFLMDSP 60  
Db 1 MKEAIKVAQVQADPIYMDLKATVDKTELMEEAARNNARLIAPETWIPGYPMFLMDSP 60  
Qy 61 AMAQPFQYQYHENSLELDGPOAKRISDAKRLGIMVTLGMSERVGTLTYSQWFIQNGD 120  
Db 61 AMAQPFQYQYHENSLELDGPOAKRISDAKRLGIMVTLGMSERVGTLTYSQWFIQNGD 120  
Qy 121 TIGARRKLKPTFVERTLFGEGDSSSLAVFETSVGRIGLCCWEHLQPLTKYALYAONEEI 180  
Db 121 TIGARRKLKPTFVERTLFGEGDSSSLAVFETSVGRIGLCCWEHLQPLTKYALYAONEEI 180  
Qy 181 HCAAMPFSIYPNAKALGPDVNVAAASRIYAVEGQCFVLASCALVSQSMIDMLCTDDEKH 240  
Db 181 HCAAMPFSIYPNAKALGPDVNVAAASRIYAVEGQCFVLASCALVSQSMIDMLCTDDEKH 240  
Qy 241 ALLIAGGHSRIIGPDGDLVAPLAENEGILYANLDPGVRIILAKMAADPAHYSRPDIT 300  
Db 241 ALLIAGGHSRIIGPDGDLVAPLAENEGILYANLDPGVRIILAKMAADPAHYSRPDIT 300  
Qy 301 RLIDRSPKLPVEIEGDLRPYALGKASETGAOLEEI 337  
Db 301 RLIDRSPKLPVEIEGDLRPYALGKASETGAOLEEI 337

RESULT 4

```
US-10-440-523-386
; Sequence 386, Application US/10440523
; Publication No. US20040014195A1
GENERAL INFORMATION:
; APPLICANT: Short, Jay
; APPLICANT: Weiner, David
; APPLICANT: Chaplin, Jennifer
; APPLICANT: Chi, Ellen
; APPLICANT: Milan, Aileen
; APPLICANT: Desantis, Grace
; APPLICANT: Madden, Mark
; APPLICANT: Burk, Mark
; TITLE OF INVENTION: Nitrites
; FILE REFERENCE: Docket No. US20040014195A1 DIV-013US
; CURRENT APPLICATION NUMBER: US/10/440,523
; PRIOR FILING DATE: 2003-05-15
; PRIOR APPLICATION NUMBER: US/10/146,772
; PRIOR FILING DATE: 2002-05-15
; PRIOR APPLICATION NUMBER: US 60/309,006
; PRIOR FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US 60/351,336
; PRIOR FILING DATE: 2002-01-22
; PRIOR APPLICATION NUMBER: US 60/300,189
; PRIOR FILING DATE: 2001-06-21
; PRIOR APPLICATION NUMBER: US 09/751,299
; PRIOR FILING DATE: 2000-12-28
; PRIOR APPLICATION NUMBER: US 60/254,414
; PRIOR FILING DATE: 2000-12-07
; PRIOR APPLICATION NUMBER: US 60/173,609
; PRIOR FILING DATE: 1999-12-29
; NUMBER OF SEQ ID NOS: 386
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 386
; LENGTH: 337
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Obtained from an environmental sample
US-10-440-523-386

Query Match          99.8%; Score 1761; DB 4; Length 337;
Best Local Similarity 99.7%; Pred. No. 3.5e-176;
Matches 336; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKEAIKVCVQAAPITMDLEATVDKTIEMEEAARNARLAFPEWTIPGYPMFLMDSP 60
DB 1 MKEAIKVCVQAAPITMDLEATVDKTIEMEEAARNARLAFPEWTIPGYPMFLMDSP 60
QY 61 AAMQFVQYHENSLELDGPOAKRISDAKRLGIMVTLGMSERVGTLTYSQWFIGDNGD 120
DB 61 AAMQFVQYHENSLELDGPOAKRISDAKRLGIMVTLGMSERVGTLTYSQWFIGDNGD 120
QY 121 TIGARRKLKPTFVERTLFGEGDSSLAVFETSVGRGLGCWEHLQPLTKYALYAONEEI 180
DB 121 TIGARRKLKPTFVERTLFGEGDSSLAVFETSVGRGLGCWEHLQPLTKYALYAONEEI 180
QY 181 HCAAMPFSLYPNAAKALGPVNVAAASRIYAVEGQCFVLASCALVSQSMIDMLCTDDEKH 240
DB 181 HCAAMPFSLYPNAAKALGPVNVAAASRIYAVEGQCFVLASCALVSQSMIDMLCTDDEKH 240
QY 241 ALLLAGGSHRIIGPDGDLVAPLAENEGILYANLDPGVRIILAKMAADPAGHSRPDIT 300
DB 241 ALLLAGGSHRIIGPDGDLVAPLAENEGILYANLDPGVRIILAKMAADPAGHSRPDIT 300
QY 301 RLIDRSPKLPVVEIEGDLRPYALGKASGTGAOLEEI 337
DB 301 RLIDRSPKLPVVEIEGDLRPYALGKASGTGAOLEEI 337

RESULT 5
US-10-440-503-386
; Sequence 386, Application US/10440503
; Publication No. US20040038419A1
```

```
GENERAL INFORMATION:
; APPLICANT: Weiner, David Paul
; APPLICANT: Chaplin, Jennifer Ann
; APPLICANT: Chi, Ellen
; APPLICANT: Milan, Aileen
; APPLICANT: Desantis, Grace
; APPLICANT: Burk, Mark J.
; APPLICANT: McQuaid, Jeffrey
; APPLICANT: Stegas, Justin
; TITLE OF INVENTION: ASSAYS AND KITS FOR DETECTING THE PRESENCE OF
; FILE REFERENCE: 09010-900001
; CURRENT APPLICATION NUMBER: US/10/440,503
; PRIOR FILING DATE: 2003-05-15
; PRIOR APPLICATION NUMBER: US 60/380,737
; PRIOR FILING DATE: 2002-05-15
; NUMBER OF SEQ ID NOS: 386
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 386
; LENGTH: 337
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Obtained from an environmental sample
US-10-440-503-386

Query Match          99.8%; Score 1761; DB 4; Length 337;
Best Local Similarity 99.7%; Pred. No. 3.5e-176;
Matches 336; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKEAIKVCVQAAPITMDLEATVDKTIEMEEAARNARLAFPEWTIPGYPMFLMDSP 60
DB 1 MKEAIKVCVQAAPITMDLEATVDKTIEMEEAARNARLAFPEWTIPGYPMFLMDSP 60
QY 61 AAMQFVQYHENSLELDGPOAKRISDAKRLGIMVTLGMSERVGTLTYSQWFIGDNGD 120
DB 61 AAMQFVQYHENSLELDGPOAKRISDAKRLGIMVTLGMSERVGTLTYSQWFIGDNGD 120
QY 121 TIGARRKLKPTFVERTLFGEGDSSLAVFETSVGRGLGCWEHLQPLTKYALYAONEEI 180
DB 121 TIGARRKLKPTFVERTLFGEGDSSLAVFETSVGRGLGCWEHLQPLTKYALYAONEEI 180
QY 181 HCAAMPFSLYPNAAKALGPVNVAAASRIYAVEGQCFVLASCALVSQSMIDMLCTDDEKH 240
DB 181 HCAAMPFSLYPNAAKALGPVNVAAASRIYAVEGQCFVLASCALVSQSMIDMLCTDDEKH 240
QY 241 ALLLAGGSHRIIGPDGDLVAPLAENEGILYANLDPGVRIILAKMAADPAGHSRPDIT 300
DB 241 ALLLAGGSHRIIGPDGDLVAPLAENEGILYANLDPGVRIILAKMAADPAGHSRPDIT 300
QY 301 RLIDRSPKLPVVEIEGDLRPYALGKASGTGAOLEEI 337
DB 301 RLIDRSPKLPVVEIEGDLRPYALGKASGTGAOLEEI 337

RESULT 6
US-10-461-925-386
; Sequence 386, Application US/10461925
; Publication No. US20040053378A1
GENERAL INFORMATION:
; APPLICANT: Mark J. Burk
; APPLICANT: Morgan, Brian
; APPLICANT: Desantis, Grace
; APPLICANT: Zhu, Zoulin
; TITLE OF INVENTION: PROCESSES FOR MAKING (R)-ETHYL 4-CYANO-3-HYDROXYBUTYRIC ACID
; FILE REFERENCE: 09010-270001
; CURRENT APPLICATION NUMBER: US/10/461,925
; PRIOR FILING DATE: 2003-06-13
; PRIOR APPLICATION NUMBER: US 60/389,317
; PRIOR FILING DATE: 2002-06-13
; PRIOR APPLICATION NUMBER: US 60/392,944
; PRIOR FILING DATE: 2002-06-28
; NUMBER OF SEQ ID NOS: 386
```

```
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 386
; LENGTH: 337
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Obtained from an environmental sample
US-10-461-925-386

Query Match          99.8%; Score 1761; DB 4; Length 337;
Best Local Similarity 99.7%; Pred. No. 3.5e-176;
Matches 336; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MKEAIKVCVQAPAPYMDLEATVDTKTIELMEBAARNARLAFPEWTIPGYPMFLMDSP 60
DB 1 MKEAIKVCVQAPAPYMDLKAIVDTKTIELMEBAARNARLAFPEWTIPGYPMFLMDSP 60
QY 61 AMAMQFVRQYHENSLELDGPQAKRISDAKRLGIMVTLGMSERVGGTLYISQWFIQNGD 120
DB 61 AMAMQFVRQYHENSLELDGPQAKRISDAKRLGIMVTLGMSERVGGTLYISQWFIQNGD 120
QY 121 TIGARRKLPFTVEVETLFGEGDSSLAVFETSVGRLGICCEHLOPLTKYALYAQNEEI 180
DB 121 TIGARRKLPFTVEVETLFGEGDSSLAVFETSVGRLGICCEHLOPLTKYALYAQNEEI 180
QY 181 HCAAMPFSFLYPRNAKALGPDVNVAAASRIYAVEGQCFVLASCALVSQSMIDMLCTDDEKH 240
DB 181 HCAAMPFSFLYPRNAKALGPDVNVAAASRIYAVEGQCFVLASCALVSQSMIDMLCTDDEKH 240
QY 241 ALLAGGHSRIIGPDGDLVAPLAENEGILYANLDPGVRLLAKMADPAGHSRPDIT 300
DB 241 ALLAGGHSRIIGPDGDLVAPLAENEGILYANLDPGVRLLAKMADPAGHSRPDIT 300
QY 301 RLIIIRSPKLPVEIEIGDLRPYALGKASSETGAQLEI 337
DB 301 RLIIIRSPKLPVEIEIGDLRPYALGKASSETGAQLEI 337

RESULT 7
US-10-146-772-48
; Sequence 48, Application US/10146772
; Publication No. US20030124698A1
; GENERAL INFORMATION:
; APPLICANT: Short, Jay
; APPLICANT: Weiner, David
; APPLICANT: Chaplin, Jennifer
; APPLICANT: Chi, Ellen
; APPLICANT: Milan, Aileen
; APPLICANT: Desantis, Mark
; APPLICANT: Madden, Mark
; TITLE OF INVENTION: Nitrlases
; FILE REFERENCE: Docket No. US20030124698A1 DIV-013US
; CURRENT APPLICATION NUMBER: US/10/146,772
; PRIOR FILING DATE: 2002-05-15
; PRIOR APPLICATION NUMBER: US 60/309,006
; PRIOR FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US 60/351,336
; PRIOR FILING DATE: 2002-01-22
; PRIOR APPLICATION NUMBER: US 60/300,189
; PRIOR FILING DATE: 2001-06-21
; PRIOR APPLICATION NUMBER: US 09/751,299
; PRIOR FILING DATE: 2000-12-28
; PRIOR APPLICATION NUMBER: US 60/254,414
; PRIOR FILING DATE: 2000-12-07
; PRIOR APPLICATION NUMBER: US 60/173,609
; PRIOR FILING DATE: 1999-12-29
; NUMBER OF SEQ ID NOS: 386
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 48
; LENGTH: 337
; TYPE: PRT
; ORGANISM: Unknown
```

```
; FEATURE:
; OTHER INFORMATION: Obtained from an environmental sample
US-10-146-772-48

Query Match          75.9%; Score 1339; DB 4; Length 337;
Best Local Similarity 75.2%; Pred. No. 9.1e-132;
Matches 252; Conservative 30; Mismatches 47; Indels 6; Gaps 1;

QY 1 MKEAIKVCVQAPAPYMDLEATVDTKTIELMEBAARNARLAFPEWTIPGYPMFLMDSP 60
DB 1 MKEAIKVCVQAPAPYMDLEATVDTKTIELMEBAARNARLAFPEWTIPGYPMFLMDSP 60
QY 61 AMAMQFVRQYHENSLELDGPQAKRISDAKRLGIMVTLGMSERVGGTLYISQWFIQNGD 120
DB 61 AMAMQFVRQYHENSLELDGPQAKRISDAKRLGIMVTLGMSERVGGTLYISQWFIQNGD 120
QY 121 TIGARRKLPFTVEVETLFGEGDSSLAVFETSVGRLGICCEHLOPLTKYALYAQNEEI 180
DB 121 TIGARRKLPFTVEVETLFGEGDSSLAVFETSVGRLGICCEHLOPLTKYALYAQNEEI 180
QY 181 HCAAMPFSFLYPRNAKALGPDVNVAAASRIYAVEGQCFVLASCALVSQSMIDMLCTDDEKH 240
DB 181 HCAAMPFSFLYPRNAKALGPDVNVAAASRIYAVEGQCFVLASCALVSQSMIDMLCTDDEKH 240
QY 241 ALLAGGHSRIIGPDGDLVAPLAENEGILYANLDPGVRLLAKMADPAGHSRPDIT 300
DB 241 ALLAGGHSRIIGPDGDLVAPLAENEGILYANLDPGVRLLAKMADPAGHSRPDIT 300
QY 301 RLIIIRSPKLPVEIEIGDLRPYALGKASSETGAQLE 335
DB 301 RLIIIRSPKLPVEIEIGDLRPYALGKASSETGAQLE 335

RESULT 8
US-10-146-772-58
; Sequence 58, Application US/10146772
; Publication No. US20030124698A1
; GENERAL INFORMATION:
; APPLICANT: Short, Jay
; APPLICANT: Weiner, David
; APPLICANT: Chaplin, Jennifer
; APPLICANT: Chi, Ellen
; APPLICANT: Milan, Aileen
; APPLICANT: Desantis, Grace
; APPLICANT: Madden, Mark
; TITLE OF INVENTION: Nitrlases
; FILE REFERENCE: Docket No. US20030124698A1 DIV-013US
; CURRENT APPLICATION NUMBER: US/10/146,772
; PRIOR FILING DATE: 2002-05-15
; PRIOR APPLICATION NUMBER: US 60/309,006
; PRIOR FILING DATE: 2001-07-30
; PRIOR APPLICATION NUMBER: US 60/351,336
; PRIOR FILING DATE: 2002-01-22
; PRIOR APPLICATION NUMBER: US 60/300,189
; PRIOR FILING DATE: 2001-06-21
; PRIOR APPLICATION NUMBER: US 09/751,299
; PRIOR FILING DATE: 2000-12-28
; PRIOR APPLICATION NUMBER: US 60/254,414
; PRIOR FILING DATE: 2000-12-07
; PRIOR APPLICATION NUMBER: US 60/173,609
; PRIOR FILING DATE: 1999-12-29
; NUMBER OF SEQ ID NOS: 386
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 58
; LENGTH: 337
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Obtained from an environmental sample
US-10-146-772-58

Query Match          75.9%; Score 1339; DB 4; Length 337;
```



Best Local Similarity 75.2%; Pred. No. 9.1e-132;  
Matches 252; Conservative 30; Mismatches 47; Indels 6; Gaps 1;

```
QY 1 MKEAIKVCQVQAPVIMDEATYDVTITELMEBAARNARLIAPETWIPGYPMFLMDSP 60
DB 1 VKEAIKVCQVQAPVFLDLDATVDKTVALLIEBAARNARLIAPETWIPGYPMFLMDSP 60
QY 61 AMAMQFVRQYHENSILDEGPQAKRISDAKRLGIWTLGMSERVGCTLYISQWFIGDND 120
DB 61 AMQMCFVRHYHENSILVDSBPQAKRISDAKRLGIWTLGMSERVGCTLYISQWFIGDND 120
QY 121 TTGARRKLPKPTVEERTLFGEGDSSLAPEVTSVGLGGLCCWEHLQPLTKYALYAONEBI 180
DB 121 TAGLRKKLPKPTVEERTLFGEGDSSLSFTDTPGLVGLGCCWEHLQPLSKYALYAONEBI 180
QY 181 HCAAMPSESLYPNAALGPDVNVVAASTRIVAVGQCFVLASCALVSQSMIDMLCTDEKH 240
DB 181 HFAAMPSESLYRQATVEVLGPENVVAASRIYAVGQCFVLASCALVSPMIEMLCCTDEKH 240
QY 241 ALLIAGGHSRIIGPDGDLVAPLAENEGILYANLDPCVRLAKMAADPAGHSRPDIT 300
DB 241 SLIAGGGYSRIIGPDGSDLPALPGENEGILYATLDPAARIYAKTAAADPAGHSRPDVT 300
QY 301 RLILNRSKLPVVEIEGDLRPYALGKASETGAOLE 335
DB 301 RLILNRSANQPVVEVGREI-----PASAQGEVE 329
```

RESULT 9  
US-10-241-742-48

/ Sequence 48, Application US/10241742

/ Publication No. US20040002147A1

/ GENERAL INFORMATION:

/ APPLICANT: Short, Jay

/ APPLICANT: Weiner, David

/ APPLICANT: Chaplin, Jennifer

/ APPLICANT: Chi, Ellen

/ APPLICANT: Milan, Aileen

/ APPLICANT: Desantis, Grace

/ APPLICANT: Madden, Mark

/ TITLE OF INVENTION: Nitrlaases

/ FILE REFERENCE: Docket No. US20040002147A1 DIV-013US

/ CURRENT APPLICATION NUMBER: US/10/241,742

/ PRIOR FILING DATE: 2002-09-09

/ PRIOR APPLICATION NUMBER: US/10/146,772

/ PRIOR FILING DATE: 2002-05-15

/ PRIOR APPLICATION NUMBER: US 60/309,006

/ PRIOR FILING DATE: 2001-07-30

/ PRIOR APPLICATION NUMBER: US 60/351,336

/ PRIOR FILING DATE: 2002-01-22

/ PRIOR APPLICATION NUMBER: US 60/300,189

/ PRIOR FILING DATE: 2001-06-21

/ PRIOR APPLICATION NUMBER: US 09/751,299

/ PRIOR FILING DATE: 2000-12-28

/ PRIOR APPLICATION NUMBER: US 60/254,414

/ PRIOR FILING DATE: 2000-12-07

/ PRIOR APPLICATION NUMBER: US 60/173,609

/ PRIOR FILING DATE: 1999-12-29

/ NUMBER OF SEQ ID NOS: 386

/ SOFTWARE: FastSeq for Windows Version 4.0

/ SEQ ID NO 48

/ LENGTH: 337

/ TYPE: PRT

/ ORGANISM: Unknown

/ FEATURE:

/ OTHER INFORMATION: Obtained from an environmental sample

US-10-241-742-48

Query Match 75.9%; Score 1339; DB 4; Length 337;  
Best Local Similarity 75.2%; Pred. No. 9.1e-132;  
Matches 252; Conservative 30; Mismatches 47; Indels 6; Gaps 1;

```
QY 1 MKEAIKVCQVQAPVIMDEATYDVTITELMEBAARNARLIAPETWIPGYPMFLMDSP 60
DB 1 VKEAIKVCQVQAPVFLDLDATVDKTVALLIEBAARNARLIAPETWIPGYPMFLMDSP 60
QY 61 AMAMQFVRQYHENSILDEGPQAKRISDAKRLGIWTLGMSERVGCTLYISQWFIGDND 120
DB 61 AMQMCFVRHYHENSILVDSBPQAKRISDAKRLGIWTLGMSERVGCTLYISQWFIGDND 120
QY 121 TTGARRKLPKPTVEERTLFGEGDSSLAPEVTSVGLGGLCCWEHLQPLTKYALYAONEBI 180
DB 121 TAGLRKKLPKPTVEERTLFGEGDSSLSFTDTPGLVGLGCCWEHLQPLSKYALYAONEBI 180
QY 181 HCAAMPSESLYPNAALGPDVNVVAASTRIVAVGQCFVLASCALVSQSMIDMLCTDEKH 240
DB 181 HFAAMPSESLYRQATVEVLGPENVVAASRIYAVGQCFVLASCALVSPMIEMLCCTDEKH 240
QY 241 ALLIAGGHSRIIGPDGDLVAPLAENEGILYANLDPCVRLAKMAADPAGHSRPDIT 300
DB 241 SLIAGGGYSRIIGPDGSDLPALPGENEGILYATLDPAARIYAKTAAADPAGHSRPDVT 300
QY 301 RLILNRSKLPVVEIEGDLRPYALGKASETGAOLE 335
DB 301 RLILNRSANQPVVEVGREI-----PASAQGEVE 329
```

RESULT 10  
US-10-241-742-58

/ Sequence 58, Application US/10241742

/ Publication No. US20040002147A1

/ GENERAL INFORMATION:

/ APPLICANT: Short, Jay

/ APPLICANT: Weiner, David

/ APPLICANT: Chaplin, Jennifer

/ APPLICANT: Chi, Ellen

/ APPLICANT: Milan, Aileen

/ APPLICANT: Desantis, Grace

/ APPLICANT: Madden, Mark

/ TITLE OF INVENTION: Nitrlaases

/ FILE REFERENCE: Docket No. US20040002147A1 DIV-013US

/ CURRENT APPLICATION NUMBER: US/10/241,742

/ PRIOR FILING DATE: 2002-09-09

/ PRIOR APPLICATION NUMBER: US/10/146,772

/ PRIOR FILING DATE: 2002-05-15

/ PRIOR APPLICATION NUMBER: US 60/309,006

/ PRIOR FILING DATE: 2001-07-30

/ PRIOR APPLICATION NUMBER: US 60/351,336

/ PRIOR FILING DATE: 2002-01-22

/ PRIOR APPLICATION NUMBER: US 60/300,189

/ PRIOR FILING DATE: 2001-06-21

/ PRIOR APPLICATION NUMBER: US 09/751,299

/ PRIOR FILING DATE: 2000-12-28

/ PRIOR APPLICATION NUMBER: US 60/254,414

/ PRIOR FILING DATE: 2000-12-07

/ PRIOR APPLICATION NUMBER: US 60/173,609

/ PRIOR FILING DATE: 1999-12-29

/ NUMBER OF SEQ ID NOS: 386

/ SOFTWARE: FastSeq for Windows Version 4.0

/ SEQ ID NO 58

/ LENGTH: 337

/ TYPE: PRT

/ ORGANISM: Unknown

/ FEATURE:

/ OTHER INFORMATION: Obtained from an environmental sample

US-10-241-742-58

Query Match 75.9%; Score 1339; DB 4; Length 337;  
Best Local Similarity 75.2%; Pred. No. 9.1e-132;  
Matches 252; Conservative 30; Mismatches 47; Indels 6; Gaps 1;

```
QY 1 MKEAIKVCQVQAPVIMDEATYDVTITELMEBAARNARLIAPETWIPGYPMFLMDSP 60
DB 1 VKEAIKVCQVQAPVFLDLDATVDKTVALLIEBAARNARLIAPETWIPGYPMFLMDSP 60
```



```
Db 121 TAGLRKRLKPTHTVETLFGEGDSSLSSTFDPLVGLGGLCCWEHLQPLSKYALYAQNEI 180
Qy 181 HCAAMPSSILVFNAAKALGPDVNVAAASRIYAVEGOCFVLASCALVSQSMIDMLCTDDEKH 240
Db 181 HFAAMPSSISYRQATEVIGPEVNVAAASRIYAVEGOCFVLASCALVSPEMIEMLCTDESKH 240
Qy 241 ALLLAGGHSRIIGPDGDLVAPLAENEGILYANLDPGVRIILAKMAADPAGHSRDPIT 300
Db 241 SLLOAGGYSRIIGPDGDLARPLGENEGILYATLDPAARIYAKTAADPAGHSRDPVT 300
Qy 301 RLILDRSPKLPVETIEGDLRPYALGKASETGAQLE 335
Db 301 RLILNRSANQPVVEVGREI-----PASAQGFVE 329

RESULT 13
US-10-440-503-48
; Sequence 48, Application US/10440503
; Publication No. US20040038419A1
; GENERAL INFORMATION:
; APPLICANT: Weiner, David Paul
; APPLICANT: Chaplin, Jennifer Ann
; APPLICANT: Chi, Ellen
; APPLICANT: Milan, Aileen
; APPLICANT: Desantis, Grace
; APPLICANT: Burk, Mark J
; APPLICANT: McQuaid, Jeffrey
; APPLICANT: Stege, Justin
; TITLE OF INVENTION: ASSAYS AND KITS FOR DETECTING THE PRESENCE OF
; FILE REFERENCE: 09010-900001
; CURRENT APPLICATION NUMBER: US/10/440,503
; PRIOR FILING DATE: 2003-05-15
; PRIOR APPLICATION NUMBER: US 60/380,737
; NUMBER OF SEQ ID NOS: 386
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 48
; LENGTH: 337
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Obtained from an environmental sample
US-10-440-503-48

Query Match 75.9%; Score 1339; DB 4; Length 337;
Best Local Similarity 75.2%; Pred. No. 9,1e-132;
Matches 252; Conservative 30; Mismatches 47; Indels 6; Gaps 1;

Qy 1 MKEAIKVCVQAAPTYMDEATVDKTIEMEAARNNAKRLAFPTWIPGYPWFLMLDSP 60
Db 1 VKEAIKVCVQAAPVFDLDATVDKTVALLIEAARNKRLAFPTWIPGYPWFLMLDSP 60
Qy 61 AMAQFVRYHENSLELDGPOAKRISDAKRLGIWTLGMSERVGGLYISQWFIQDNGD 120
Db 61 AMQFVRYHENSLELDGPOAKRISDAKRLGIWTLGMSERVGGLYISQWFIQDNGD 120
Qy 61 AMQFVRYHENSLELDGPOAKRISDAKRLGIWTLGMSERVGGLYISQWFIQDNGD 120
Db 61 AMQFVRYHENSLELDGPOAKRISDAKRLGIWTLGMSERVGGLYISQWFIQDNGD 120
Qy 121 TIGARRKRLKPTHTVETLFGEGDSSLSSTFDPLVGLGGLCCWEHLQPLSKYALYAQNEI 180
Db 121 TAGLRKRLKPTHTVETLFGEGDSSLSSTFDPLVGLGGLCCWEHLQPLSKYALYAQNEI 180
Qy 181 HCAAMPSSILVFNAAKALGPDVNVAAASRIYAVEGOCFVLASCALVSQSMIDMLCTDDEKH 240
Db 181 HFAAMPSSISYRQATEVIGPEVNVAAASRIYAVEGOCFVLASCALVSPEMIEMLCTDESKH 240
Qy 241 ALLLAGGHSRIIGPDGDLVAPLAENEGILYANLDPGVRIILAKMAADPAGHSRDPIT 300
Db 241 SLLOAGGYSRIIGPDGDLARPLGENEGILYATLDPAARIYAKTAADPAGHSRDPVT 300
Qy 301 RLILDRSPKLPVETIEGDLRPYALGKASETGAQLE 335
Db 301 RLILNRSANQPVVEVGREI-----PASAQGFVE 329
```

```
RESULT 14
US-10-440-503-58
; Sequence 58, Application US/10440503
; Publication No. US20040038419A1
; GENERAL INFORMATION:
; APPLICANT: Weiner, David Paul
; APPLICANT: Chaplin, Jennifer Ann
; APPLICANT: Chi, Ellen
; APPLICANT: Milan, Aileen
; APPLICANT: Desantis, Grace
; APPLICANT: Burk, Mark J
; APPLICANT: McQuaid, Jeffrey
; APPLICANT: Stege, Justin
; TITLE OF INVENTION: ASSAYS AND KITS FOR DETECTING THE PRESENCE OF
; FILE REFERENCE: 09010-900001
; CURRENT APPLICATION NUMBER: US/10/440,503
; PRIOR FILING DATE: 2003-05-15
; PRIOR APPLICATION NUMBER: US 60/380,737
; NUMBER OF SEQ ID NOS: 386
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 58
; LENGTH: 337
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Obtained from an environmental sample
US-10-440-503-58
```

```
Query Match 75.9%; Score 1339; DB 4; Length 337;
Best Local Similarity 75.2%; Pred. No. 9,1e-132;
Matches 252; Conservative 30; Mismatches 47; Indels 6; Gaps 1;

Qy 1 MKEAIKVCVQAAPTYMDEATVDKTIEMEAARNNAKRLAFPTWIPGYPWFLMLDSP 60
Db 1 VKEAIKVCVQAAPVFDLDATVDKTVALLIEAARNKRLAFPTWIPGYPWFLMLDSP 60
Qy 61 AMAQFVRYHENSLELDGPOAKRISDAKRLGIWTLGMSERVGGLYISQWFIQDNGD 120
Db 61 AMQFVRYHENSLELDGPOAKRISDAKRLGIWTLGMSERVGGLYISQWFIQDNGD 120
Qy 61 AMQFVRYHENSLELDGPOAKRISDAKRLGIWTLGMSERVGGLYISQWFIQDNGD 120
Db 61 AMQFVRYHENSLELDGPOAKRISDAKRLGIWTLGMSERVGGLYISQWFIQDNGD 120
Qy 121 TIGARRKRLKPTHTVETLFGEGDSSLSSTFDPLVGLGGLCCWEHLQPLSKYALYAQNEI 180
Db 121 TAGLRKRLKPTHTVETLFGEGDSSLSSTFDPLVGLGGLCCWEHLQPLSKYALYAQNEI 180
Qy 181 HCAAMPSSILVFNAAKALGPDVNVAAASRIYAVEGOCFVLASCALVSQSMIDMLCTDDEKH 240
Db 181 HFAAMPSSISYRQATEVIGPEVNVAAASRIYAVEGOCFVLASCALVSPEMIEMLCTDESKH 240
Qy 241 ALLLAGGHSRIIGPDGDLVAPLAENEGILYANLDPGVRIILAKMAADPAGHSRDPIT 300
Db 241 SLLOAGGYSRIIGPDGDLARPLGENEGILYATLDPAARIYAKTAADPAGHSRDPVT 300
Qy 301 RLILDRSPKLPVETIEGDLRPYALGKASETGAQLE 335
Db 301 RLILNRSANQPVVEVGREI-----PASAQGFVE 329

RESULT 15
US-10-461-925-48
; Sequence 48, Application US/10461925
; Publication No. US20040053378A1
; GENERAL INFORMATION:
; APPLICANT: Mark J. Burk
; APPLICANT: Desantis, Grace
; APPLICANT: Morgan, Brian
; APPLICANT: Zhu, Zoulin
; TITLE OF INVENTION: PROCESSES FOR MAKING (R)-ETHYL 4-CYANO-3-HYDROXYBUTYRIC ACID
; FILE REFERENCE: 09010-270001
; CURRENT APPLICATION NUMBER: US/10/461,925
```

```

; CURRENT FILING DATE: 2003-06-13
; PRIOR APPLICATION NUMBER: US 60/389,317
; PRIOR FILING DATE: 2002-06-13
; PRIOR APPLICATION NUMBER: US 60/392,944
; PRIOR FILING DATE: 2002-06-28
; NUMBER OF SEQ ID NOS: 386
; SOFTWARE: FASTSEQ for Windows Version 4.0
; SEQ ID NO 48
; LENGTH: 337
; TYPE: PRT
; ORGANISM: Unknown
; FEATURE:
; OTHER INFORMATION: Obtained from an environmental sample
US-10-461-925-48

```

```

Query Match 75.9%; Score 1339; DB 4; Length 337;
Best Local Similarity 75.2%; Pred. No. 9,1e-132;
Matches 252; Conservative 30; Mismatches 47; Indels 6; Gaps 1;

```

```

QY 1 MKEAIKVCVQAAPYMDLEATVDKTIEMEEAARNNAALIAPEPTWIPGYPMFLMDSP 60
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 1 VKEAIKVCVQAAPYMDLEATVDKTIEMEEAARNNAALIAPEPTWIPGYPMFLMDSP 60
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 61 AWAMQFVRQYHENSLELDGPQAKRISDAKRLGIMVTLGMSERYGCTIYISQWFIGDNGD 120
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 61 AWAMQFVRQYHENSLELDGPQAKRISDAKRLGIMVTLGMSERYGCTIYISQWFIGDNGD 120
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 121 TIGARRKLPKPTFVERTLFGEGDSSLAVENTSVGRIGLCCWEHLQPLTKYALYAONEEI 180
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 121 TIGARRKLPKPTFVERTLFGEGDSSLAVENTSVGRIGLCCWEHLQPLTKYALYAONEEI 180
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 181 HCAAMPSESLYPNAKALGPDVNAASRIYAVEGQCFVLASGALVYSQSMIDMLCTDDEKH 240
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 181 HCAAMPSESLYPNAKALGPDVNAASRIYAVEGQCFVLASGALVYSQSMIDMLCTDDEKH 240
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 241 ALLLAGGHSRIIGPDGDLVAPLAENERGIUYANIDPGVRIIAXMAADPAGHSRPDIT 300
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 241 ALLLAGGHSRIIGPDGDLVAPLAENERGIUYANIDPGVRIIAXMAADPAGHSRPDIT 300
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 301 RLILDRSPKLPVVEIEGDLRPYALGKASETGAOLE 335
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 301 RLILDRSPKLPVVEIEGDLRPYALGKASETGAOLE 335
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
QY 301 RLILNRSANQPVVEVGREI-----PASAQFEVE 329
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
Db 301 RLILNRSANQPVVEVGREI-----PASAQFEVE 329
   :|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:

```

Search completed: April 27, 2006, 01:06:08  
 Job time : 109.044 secs

GenCore version 5.1.7  
Copyright (c) 1993 - 2006 Bioacceleration Ltd.

OM protein - protein search, using sw model

Run on: April 27, 2006, 01:02:36 ; Search time 17.7628 Seconds  
(without alignments)  
862.816 Million cell updates/sec

Title: US-09-751-299-4

Perfect score: 1765  
Sequence: 1 MKEAIKVCQAAPLYMDLE.....DLRPVALGKASGTGAQLEEI 337

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 232119 seqs, 4547862 residues

Total number of hits satisfying chosen parameters: 232119

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database :

Published Applications AA New:\*  
1: /SIDS/prodata/2/pubppa/US08\_NEW\_PUB.pep:\*  
2: /SIDS/prodata/2/pubppa/US06\_NEW\_PUB.pep:\*  
3: /SIDS/prodata/2/pubppa/US07\_NEW\_PUB.pep:\*  
4: /SIDS/prodata/2/pubppa/US09\_NEW\_PUB.pep:\*  
5: /SIDS/prodata/2/pubppa/US10\_NEW\_PUB.pep:\*  
6: /SIDS/prodata/2/pubppa/US11\_NEW\_PUB.pep:\*  
7: /SIDS/prodata/2/pubppa/US60\_NEW\_PUB.pep:\*  
8: /SIDS/prodata/2/pubppa/US60\_NEW\_PUB.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	822	46.6	356	US-10-537-075-7	Sequence 7, Appli
2	780	44.2	369	US-10-919-182-18	Sequence 18, Appli
3	777	44.0	369	US-10-919-182-16	Sequence 16, Appli
4	776	44.0	369	US-10-919-182-4	Sequence 4, Appli
5	776	44.0	369	US-10-919-182-6	Sequence 6, Appli
6	774	43.9	369	US-10-919-182-12	Sequence 12, Appli
7	774	43.9	369	US-10-919-182-14	Sequence 14, Appli
8	765	43.3	369	US-10-919-182-8	Sequence 8, Appli
9	451	25.6	333	US-11-096-568A-20687	Sequence 20687, A
10	451	25.6	351	US-11-096-568A-20686	Sequence 20686, A
11	446.5	25.3	330	US-11-096-568A-6955	Sequence 6955, Ap
12	446.5	25.3	350	US-11-096-568A-6954	Sequence 6954, Ap
13	423	18.3	233	US-11-096-568A-6956	Sequence 6956, Ap
14	274.5	15.6	193	US-11-096-568A-20688	Sequence 20688, A
15	125	7.1	295	US-11-079-463-5457	Sequence 5457, Ap
16	114.5	6.5	259	US-11-045-004-822	Sequence 822, App
17	109	6.2	296	US-11-045-004-2100	Sequence 2100, Ap
18	106	6.0	460	US-11-087-099-8682	Sequence 8682, Ap
19	97	5.5	388	US-10-497-135-19	Sequence 19, Appli
20	97	5.5	388	US-11-269-215-19	Sequence 19, Appli
21	95	5.4	252	US-11-055-822-134	Sequence 134, App
22	95	5.4	252	US-11-055-822-1120	Sequence 1120, Ap
23	92	5.2	153	US-10-986-501-173	Sequence 173, App
24	92	5.2	386	US-10-497-135-20	Sequence 20, Appli
25	92	5.2	386	US-11-269-215-20	Sequence 20, Appli

26	88	5.0	463	7	US-11-087-099-5333	Sequence 5333, Ap
27	88	5.0	489	7	US-11-188-298-1840	Sequence 1840, Ap
28	86	4.9	358	7	US-11-087-099-5821	Sequence 5821, Ap
29	86	4.9	358	7	US-11-188-298-1634	Sequence 1634, A
30	85	4.8	487	7	US-11-188-298-12261	Sequence 12261, A
31	84	4.8	304	6	US-10-467-657-7410	Sequence 7410, A
32	84	4.8	369	7	US-11-188-298-15842	Sequence 15842, A
33	83	4.7	451	7	US-11-188-298-21140	Sequence 21140, A
34	82.5	4.7	3655	7	US-11-075-185-5	Sequence 5, Appli
35	82	4.6	356	7	US-11-188-298-8290	Sequence 8290, Ap
36	81.5	4.6	509	7	US-11-188-298-4316	Sequence 4316, Ap
37	81.5	4.6	1084	7	US-11-079-463-5327	Sequence 5327, Ap
38	81	4.6	455	7	US-11-188-298-11004	Sequence 11004, A
39	81	4.6	500	7	US-11-188-298-6835	Sequence 6835, Ap
40	81	4.6	585	7	US-11-045-004-1585	Sequence 1585, Ap
41	81	4.6	1065	6	US-10-511-989-28	Sequence 28, Appli
42	80.5	4.6	290	7	US-11-188-298-1550	Sequence 1550, Ap
43	80.5	4.6	306	7	US-11-188-298-17485	Sequence 17485, A
44	80.5	4.6	474	7	US-11-188-298-2797	Sequence 2797, Ap
45	80.5	4.6	475	7	US-11-087-099-2233	Sequence 2233, Ap

ALIGNMENTS

```
RESULT 1
US-10-537-075-7
; Sequence 7, Application US/10537075
; Publication No. US20060014291A1
; GENERAL INFORMATION:
; APPLICANT: Kessel, Maria
; APPLICANT: Haer, Bernhard
; TITLE OF INVENTION: L-RHAMNOSYL-INDUCIBLE EXPRESSION SYSTEMS
; FILE REFERENCE: 12810-00091-US
; CURRENT APPLICATION NUMBER: US/10/537, 075
; CURRENT FILING DATE: 2005-06-01
; PRIOR APPLICATION NUMBER: PCT/EP2003/013367
; PRIOR FILING DATE: 2003-11-27
; PRIOR APPLICATION NUMBER: DE 102 56 381.0
; PRIOR FILING DATE: 2002-12-02
; NUMBER OF SEQ ID NOS: 19
; SOFTWARE: PatentIn version 3.3
; SEQ ID NO 7
; LENGTH: 356
; TYPE: PRT
; ORGANISM: Alcaligenes faecalis
US-10-537-075-7
```

```
Query Match 46.6%; Score 822; DB 6; Length 356;
Best Local Similarity 48.2%; Pred. No. 2.8e-71;
Matches 160; Conservative 56; Mismatches 114; Indels 2; Gaps 1;

QY 2 KEAIKVCQAAPLYMDLEATVDTKTEILMEBAARNNAITAFPTWIPGFWFLDSPA 61
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 4 KRIYAAVQAASPRYDILATGVDTKTEILARQARDEGCLYFGETMLGCPHHWLGAPA 63
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 62 WAMQFVROYHNSLELDGPOAKRISDAKRLGIVNTLMSRVRGCTIYISQWFIQDNGDT 121
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 64 WSLKVSANRYANSLSDBAEFORIAQAARTLGIFALGYSSRSRGGSLYIGCLIDKDEM 123
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 122 IGARKRLKPTVVERTLFEBSGSSSLAVETSVGRUGLCCWEHLLOPLTKYALXONEI 181
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 124 LMSRRKLPTEVERTVFEBSGARLDLYSDTELGVRGALCCWEHLSPLSKVALYSQHEA 183
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 182 CAAMPSPFLYNNAKALGPDVNVASRYAYEGOCFVLASGALYSQSWIDMLCTDDEKHA 241
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 184 IAAFPSSLYEQAHALSAKYNMAASQIYSEGCFTTAAASVTQETLDMLEVEHNAP 243
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|

QY 242 LLLAGGSHRIIGPDGDLVAPLAENEGILYANLDPGVRIAKMAADPAGHYSPDITR 301
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 244 LLKVGGSMSLFAPDGRTIAPYLPFHDAAGLLIADLNMEIAFAKAINDPVGHYSKPA 303
   ::::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
```

QY 302 LLDSPKLPVVEIEGDLRPYALGKASGTGAQ 333  
DB 304 LVLDLGHRRDPWTRVHS--KSVTRREADEQGVQ 333

## RESULT 2

US-10-919-182-18  
Sequence 18, Application US/10919182  
Publication No. US20060035352A1  
GENERAL INFORMATION:

APPLICANT: E.I. duPont de Nemours and Company, Inc.

APPLICANT: Di Cosimo, Robert

APPLICANT: Payne, Mark

APPLICANT: O'Keefe, Daniel

TITLE OF INVENTION: PRODUCTION OF 3-HYDROXYCARBOXYLIC ACID USING NITRILASE MUTANTS

FILE REFERENCE: CL2584 US NA

CURRENT APPLICATION NUMBER: US/10/919,182

CURRENT FILING DATE: 2004-08-16

NUMBER OF SEQ ID NOS: 18

SOFTWARE: Patentin version 3.2

SEQ ID NO 18

LENGTH: 369

TYPE: PRT

ORGANISM: Artificial sequence

FEATURE:  
OTHER INFORMATION: Mutant nitrilase with Phe168 to Leu change

US-10-919-182-18

Query Match 44.2%; Score 780; DB 6; Length 369;  
Best Local Similarity 47.9%; Pred. No. 3.2e-67;

Matches 151; Conservative 61; Mismatches 99; Indels 4; Gaps 3;

QY 8 ACVQAPLYMDLEAVVDKTIEMEEAARNNAALIAFPETWIPGYWFMILDSPAMAMQFV 67  
DB 11 ATVOAPEVWLDADATIDKSIGIEBAQKASLIAPFEVFIPIGYWAMLDGVKXISLFT 70  
QY 68 ROYHNSLELDGPQAKRISDAKRLGIMVTLGMSRVGTLTYSQWFIQDNGDTIGARRK 127  
DB 71 SRYHNSLELDGDRMRRLQLAARNNKIALVMGYSEREGSRYSQVFIIDERGEIVANRRK 130  
QY 128 LKPTFVERTLFGEGDSSLAIVETSVGRIGLCCWEHLOPLTKYALYAONEIEHCAMPS 187  
DB 131 LKPTFVERTIYGGNGTDFLTDFPAFGRVGLNCWEHLOPLSKPMYSLGEOVHVASWPA 190  
QY 188 FS-LYENAKALGPVNVAASTRIVAVEGOCFYLAACALVSQSMIDMLCTDDEKHALLLAG 246  
DB 191 MSPLOQDPVFO-LSTEANATVTSYAIEGOTFVLCSQVIGPAIEFICLNDQORALLPOG 249  
QY 247 GGHRSRIIGDPDGDVLAFLAENEGIIYANLDPGVRILAMAADPAGHSRPPDITLLDR 306  
DB 250 CGWARIYGPDSGLAKPLAEDAGILYAEIDLEQILAKAGADPVGHYSRPDVLVSQFDP 309  
QY 307 SPKLVP--VEIEGDL 319  
DB 310 RNHTPVHRIIGIDRL 324

## RESULT 3

US-10-919-182-16  
Sequence 16, Application US/10919182  
Publication No. US20060035352A1  
GENERAL INFORMATION:

APPLICANT: E.I. duPont de Nemours and Company, Inc.

APPLICANT: Di Cosimo, Robert

APPLICANT: Payne, Mark

APPLICANT: O'Keefe, Daniel

TITLE OF INVENTION: PRODUCTION OF 3-HYDROXYCARBOXYLIC ACID USING NITRILASE MUTANTS

FILE REFERENCE: CL2584 US NA

CURRENT APPLICATION NUMBER: US/10/919,182

CURRENT FILING DATE: 2004-08-16

NUMBER OF SEQ ID NOS: 18

SOFTWARE: Patentin version 3.2

SEQ ID NO 16

LENGTH: 369  
TYPE: PRT  
ORGANISM: Artificial sequence  
FEATURE:  
OTHER INFORMATION: Mutant nitrilase with Phe168 to Val change  
US-10-919-182-16

Query Match 44.0%; Score 777; DB 6; Length 369;  
Best Local Similarity 47.6%; Pred. No. 6.3e-67;  
Matches 150; Conservative 62; Mismatches 99; Indels 4; Gaps 3;

QY 8 ACVQAPLYMDLEAVVDKTIEMEEAARNNAALIAFPETWIPGYWFMILDSPAMAMQFV 67  
DB 11 ATVOAPEVWLDADATIDKSIGIEBAQKASLIAPFEVFIPIGYWAMLDGVKXISLFT 70  
QY 68 ROYHNSLELDGPQAKRISDAKRLGIMVTLGMSRVGTLTYSQWFIQDNGDTIGARRK 127  
DB 71 SRYHNSLELDGDRMRRLQLAARNNKIALVMGYSEREGSRYSQVFIIDERGEIVANRRK 130  
QY 128 LKPTFVERTLFGEGDSSLAIVETSVGRIGLCCWEHLOPLTKYALYAONEIEHCAMPS 187  
DB 131 LKPTFVERTIYGGNGTDFLTDFPAFGRVGLNCWEHLOPLSKPMYSLGEOVHVASWPA 190  
QY 188 FS-LYENAKALGPVNVAASTRIVAVEGOCFYLAACALVSQSMIDMLCTDDEKHALLLAG 246  
DB 191 MSPLOQDPVFO-LSTEANATVTSYAIEGOTFVLCSQVIGPAIEFICLNDQORALLPOG 249  
QY 247 GGHRSRIIGDPDGDVLAFLAENEGIIYANLDPGVRILAMAADPAGHSRPPDITLLDR 306  
DB 250 CGWARIYGPDSGLAKPLAEDAGILYAEIDLEQILAKAGADPVGHYSRPDVLVSQFDP 309  
QY 307 SPKLVP--VEIEGDL 319  
DB 310 RNHTPVHRIIGIDRL 324

## RESULT 4

US-10-919-182-4  
Sequence 4, Application US/10919182  
Publication No. US20060035352A1  
GENERAL INFORMATION:

APPLICANT: E.I. duPont de Nemours and Company, Inc.

APPLICANT: Di Cosimo, Robert

APPLICANT: Payne, Mark

APPLICANT: O'Keefe, Daniel

TITLE OF INVENTION: PRODUCTION OF 3-HYDROXYCARBOXYLIC ACID USING NITRILASE MUTANTS

FILE REFERENCE: CL2584 US NA

CURRENT APPLICATION NUMBER: US/10/919,182

CURRENT FILING DATE: 2004-08-16

NUMBER OF SEQ ID NOS: 18

SOFTWARE: Patentin version 3.2

SEQ ID NO 4

LENGTH: 369

TYPE: PRT

ORGANISM: Acidovorax facilis 72W

US-10-919-182-4

Query Match 44.0%; Score 776; DB 6; Length 369;  
Best Local Similarity 47.6%; Pred. No. 7.8e-67;

Matches 150; Conservative 61; Mismatches 100; Indels 4; Gaps 3;

QY 8 ACVQAPLYMDLEAVVDKTIEMEEAARNNAALIAFPETWIPGYWFMILDSPAMAMQFV 67  
DB 11 ATVOAPEVWLDADATIDKSIGIEBAQKASLIAPFEVFIPIGYWAMLDGVKXISLFT 70  
QY 68 ROYHNSLELDGPQAKRISDAKRLGIMVTLGMSRVGTLTYSQWFIQDNGDTIGARRK 127  
DB 71 SRYHNSLELDGDRMRRLQLAARNNKIALVMGYSEREGSRYSQVFIIDERGEIVANRRK 130  
QY 128 LKPTFVERTLFGEGDSSLAIVETSVGRIGLCCWEHLOPLTKYALYAONEIEHCAMPS 187  
DB 131 LKPTFVERTIYGGNGTDFLTDFPAFGRVGLNCWEHLOPLSKPMYSLGEOVHVASWPA 190

Qy 188 FS-LYNNAAKALGPDVNVAAASRIYAVEGQCFVLASCAVSGSMIDMLCTDDEKALLLAG 246  
Db 191 MSPLOPDVFO-LSIEANATVTSVAIEGQTFVLCSTQVIGPSAIEFTFCINDEQALLPOG 249  
Qy 247 GGHSHIIPDGGDVLAPLAENEGILYANLDPGVRIILAKMADDPAGHSRDPITLLIDR 306  
Db 250 CGMAIYGPDSSELAFLAEDAEGLVIAEIDLEQIILAKAGADPVGHYSRDPVLSVQDPD 309  
Qy 307 SPKLPV--VEIEGDL 319  
Db 310 RNHTPVHRIIGIDRL 324

RESULT 5  
US-10-919-182-6  
Sequence 6, Application US/10919182  
Publication No. US20060035352A1  
GENERAL INFORMATION:  
APPLICANT: E.I. duPont de Nemours and Company, Inc.  
APPLICANT: Di Cosimo, Robert  
APPLICANT: Payne, Mark  
APPLICANT: O'Keefe, Daniel  
TITLE OF INVENTION: PRODUCTION OF 3-HYDROXYCARBOXYLIC ACID USING NITRILASE MUTANTS  
FILE REFERENCE: CL2584 US NA  
CURRENT APPLICATION NUMBER: US/10/919,182  
CURRENT FILING DATE: 2004-08-16  
NUMBER OF SEQ ID NOS: 18  
SOFTWARE: PatentIn version 3.2  
SEQ ID NO 6  
LENGTH: 369  
TYPE: PRT  
ORGANISM: Artificial sequence  
FEATURE:  
OTHER INFORMATION: Mutant nitrilase B2 and H9  
US-10-919-182-6

Query Match 44.0%; Score 776; DB 6; Length 369;  
Best Local Similarity 47.6%; Pred. No. 7.8e-67;  
Matches 150; Conservative 61; Mismatches 100; Indels 4; Gaps 3;

Qy 8 ACVQAAPIYMDLEATVDTKTIEMEEAARNNAALIAFPETWIPGYWFLMDSPPAMQFV 67  
Db 11 ATVQAEPPWMLDADATIDKSIIGIEBAQKASLIAPFEVFIPIGYWFLMDSPPAMQFV 70  
Qy 68 ROYHENSLELDGPOAKRISDAKRLGIWTLGMSERVGTLTYSQWFTGNDGDTIGARRK 127  
Db 71 SRYHENSLELDGDDRRRLQLAARRNKIALVMGYSEREAGSRYSLSQVFIIDERGEIYANRRK 130  
Qy 128 LKPTFVERTLFGEGDSSLAIVETSVERGLGCLCEHLOPLTKYALYQNEBEIHCAWPS 187  
Db 131 LKPTFVERTTIGEGNGTFLTHDPAFGRVGLNCHEHQPUSKFMWYSLGEOVHVASWPA 190  
Qy 188 FS-LYNNAAKALGPDVNVAAASRIYAVEGQCFVLASCAVSGSMIDMLCTDDEKALLLAG 246  
Db 191 MSPLOPDVFO-LSIEANATVTSVAIEGQTFVLCSTQVIGPSAIEFTFCINDEQALLPOG 249  
Qy 247 GGHSHIIPDGGDVLAPLAENEGILYANLDPGVRIILAKMADDPAGHSRDPITLLIDR 306  
Db 250 CGMAIYGPDSSELAFLAEDAEGLVIAEIDLEQIILAKAGADPVGHYSRDPVLSVQDPD 309  
Qy 307 SPKLPV--VEIEGDL 319  
Db 310 RNHTPVHRIIGIDRL 324

RESULT 6  
US-10-919-182-12  
Sequence 12, Application US/10919182  
Publication No. US20060035352A1  
GENERAL INFORMATION:  
APPLICANT: E.I. duPont de Nemours and Company, Inc.  
APPLICANT: Di Cosimo, Robert  
APPLICANT: Payne, Mark

APPLICANT: O'Keefe, Daniel  
TITLE OF INVENTION: PRODUCTION OF 3-HYDROXYCARBOXYLIC ACID USING NITRILASE MUTANTS  
FILE REFERENCE: CL2584 US NA  
CURRENT APPLICATION NUMBER: US/10/919,182  
CURRENT FILING DATE: 2004-08-16  
NUMBER OF SEQ ID NOS: 18  
SOFTWARE: PatentIn version 3.2  
SEQ ID NO 12  
LENGTH: 369  
TYPE: PRT  
ORGANISM: Artificial sequence  
FEATURE:  
OTHER INFORMATION: Mutant nitrilase with Thr210 to Cys change  
US-10-919-182-12

Query Match 43.9%; Score 774; DB 6; Length 369;  
Best Local Similarity 47.6%; Pred. No. 1.2e-66;  
Matches 150; Conservative 60; Mismatches 101; Indels 4; Gaps 3;

Qy 8 ACVQAAPIYMDLEATVDTKTIEMEEAARNNAALIAFPETWIPGYWFLMDSPPAMQFV 67  
Db 11 ATVQAEPPWMLDADATIDKSIIGIEBAQKASLIAPFEVFIPIGYWFLMDSPPAMQFV 70  
Qy 68 ROYHENSLELDGPOAKRISDAKRLGIWTLGMSERVGTLTYSQWFTGNDGDTIGARRK 127  
Db 71 SRYHENSLELDGDDRRRLQLAARRNKIALVMGYSEREAGSRYSLSQVFIIDERGEIYANRRK 130  
Qy 128 LKPTFVERTLFGEGDSSLAIVETSVERGLGCLCEHLOPLTKYALYQNEBEIHCAWPS 187  
Db 131 LKPTFVERTTIGEGNGTFLTHDPAFGRVGLNCHEHQPUSKFMWYSLGEOVHVASWPA 190  
Qy 188 FS-LYNNAAKALGPDVNVAAASRIYAVEGQCFVLASCAVSGSMIDMLCTDDEKALLLAG 246  
Db 191 MSPLOPDVFO-LSIEANATVTSVAIEGQTFVLCSTQVIGPSAIEFTFCINDEQALLPOG 249  
Qy 247 GGHSHIIPDGGDVLAPLAENEGILYANLDPGVRIILAKMADDPAGHSRDPITLLIDR 306  
Db 250 CGMAIYGPDSSELAFLAEDAEGLVIAEIDLEQIILAKAGADPVGHYSRDPVLSVQDPD 309  
Qy 307 SPKLPV--VEIEGDL 319  
Db 310 RNHTPVHRIIGIDRL 324

RESULT 7  
US-10-919-182-14  
Sequence 14, Application US/10919182  
Publication No. US20060035352A1  
GENERAL INFORMATION:  
APPLICANT: E.I. duPont de Nemours and Company, Inc.  
APPLICANT: Di Cosimo, Robert  
APPLICANT: Payne, Mark  
APPLICANT: O'Keefe, Daniel  
TITLE OF INVENTION: PRODUCTION OF 3-HYDROXYCARBOXYLIC ACID USING NITRILASE MUTANTS  
FILE REFERENCE: CL2584 US NA  
CURRENT APPLICATION NUMBER: US/10/919,182  
CURRENT FILING DATE: 2004-08-16  
NUMBER OF SEQ ID NOS: 18  
SOFTWARE: PatentIn version 3.2  
SEQ ID NO 14  
LENGTH: 369  
TYPE: PRT  
ORGANISM: Artificial sequence  
FEATURE:  
OTHER INFORMATION: Mutant nitrilase with Phe168 to Lys change  
US-10-919-182-14

Query Match 43.9%; Score 774; DB 6; Length 369;  
Best Local Similarity 47.6%; Pred. No. 1.2e-66;  
Matches 150; Conservative 61; Mismatches 100; Indels 4; Gaps 3;

Qy 8 ACVQAAPIYMDLEATVDTKTIEMEEAARNNAALIAFPETWIPGYWFLMDSPPAMQFV 67  
Db 11 ATVQAEPPWMLDADATIDKSIIGIEBAQKASLIAPFEVFIPIGYWFLMDSPPAMQFV 70  
Qy 68 ROYHENSLELDGPOAKRISDAKRLGIWTLGMSERVGTLTYSQWFTGNDGDTIGARRK 127  
Db 71 SRYHENSLELDGDDRRRLQLAARRNKIALVMGYSEREAGSRYSLSQVFIIDERGEIYANRRK 130  
Qy 128 LKPTFVERTLFGEGDSSLAIVETSVERGLGCLCEHLOPLTKYALYQNEBEIHCAWPS 187  
Db 131 LKPTFVERTTIGEGNGTFLTHDPAFGRVGLNCHEHQPUSKFMWYSLGEOVHVASWPA 190  
Qy 188 FS-LYNNAAKALGPDVNVAAASRIYAVEGQCFVLASCAVSGSMIDMLCTDDEKALLLAG 246  
Db 191 MSPLOPDVFO-LSIEANATVTSVAIEGQTFVLCSTQVIGPSAIEFTFCINDEQALLPOG 249  
Qy 247 GGHSHIIPDGGDVLAPLAENEGILYANLDPGVRIILAKMADDPAGHSRDPITLLIDR 306  
Db 250 CGMAIYGPDSSELAFLAEDAEGLVIAEIDLEQIILAKAGADPVGHYSRDPVLSVQDPD 309  
Qy 307 SPKLPV--VEIEGDL 319  
Db 310 RNHTPVHRIIGIDRL 324

```

Db      11 ATVOAEFVWMLDADATITKSGIIEBAQKASLIAPEVEIPGYPYAMIGDVKYSLSFT 70
Qy      68 ROYHENSLELDGPOAKRISDAKRLGIMVTLGMSERVGTLYISQWFIGNDGTIGARRK 127
       71 SRHYENSLELDGDMRRLQIAARANKIALVMVYSERBAGRYLSQVIERGERIVARRK 130
Qy      128 LKPTFVERTLFGEDSSSLAVFESVRLGGLCWEHLQPLTKYALYAQNEBIHCAMPS 187
       131 LKPTFVERTLFGEDSSSLAVFESVRLGGLCWEHLQPLTKYALYAQNEBIHCAMPS 190
Qy      188 FS-LYNAAKALGPDVNVVAASRIYAVEGQCFVLASCAVSSQSMIDMLCTDDEKHALLAG 246
       191 MSPAQPDVQ-LSIEANATVRSYALISGQPFVLCSTQVIGPSALETFCUNDEGRALLPQG 249
Db      247 GGHRRITGPDGDLVAPLAENEGILYANLDPGVRIILAKMAADPAHYSRPDITRLIDR 306
       250 CGMARIVGPDGSELAKPLAEDAEGLYAEIDLEQILAKAGADPVGHYSRPDVLVSQFDP 309
Qy      307 SPKLPV--VEIEGDL 319
       310 RNHTPVHRIIGIDRL 324
Db

```

```

RESULT 8
US-10-919-182-8
; Sequence 8, Application US/10919182
; Publication No. US2006003552A1
; GENERAL INFORMATION:
; APPLICANT: E.I. duPont de Nemours and Company, Inc.
; APPLICANT: Di Cosimo, Robert
; APPLICANT: Payne, Mark
; APPLICANT: O'Keefe, Daniel
; TITLE OF INVENTION: PRODUCTION OF 3-HYDROXYCARBOXYLIC ACID USING NITRILASE MUTANTS
; FILE REFERENCE: C12584 US NA
; CURRENT APPLICATION NUMBER: US/10/919,182
; CURRENT FILING DATE: 2004-08-16
; NUMBER OF SEQ ID NOS: 18
; SOFTWARE: PatentIn version 3.2
; SEQ ID NO 8
; LENGTH: 369
; TYPE: PRT
; ORGANISM: Artificial sequence
; FEATURE:
; OTHER INFORMATION: Mutant nitrilase B4
US-10-919-182-8

```

```

Query Match      43.3%; Score 765; DB 6; Length 369;
Best Local Similarity 47.6%; Pred. No. 9e-66;
Matches 150; Conservative 58; Mismatches 103; Indels 4; Gaps 3;

Qy      8 ACVOAAPIYMDLEATVDTKTIEMEEAARNNAFLIAPETWIPGYWFLMDSPAMQFV 67
       11 ATVOAEFVWMLDADATITKSGIIEBAQKASLIAPEVEIPGYPYAMIGDVKYSLSFT 70
Db      68 ROYHENSLELDGPOAKRISDAKRLGIMVTLGMSERVGTLYISQWFIGNDGTIGARRK 127
       71 SRHYENSLELDGDMRRLQIAARANKIALVMVYSERBAGRYLSQVIERGERIVARRK 130
Qy      128 LKPTFVERTLFGEDSSSLAVFESVRLGGLCWEHLQPLTKYALYAQNEBIHCAMPS 187
       131 LKPTFVERTLFGEDSSSLAVFESVRLGGLCWEHLQPLTKYALYAQNEBIHCAMPS 190
Db      188 FS-LYNAAKALGPDVNVVAASRIYAVEGQCFVLASCAVSSQSMIDMLCTDDEKHALLAG 246
       191 MSPAQPDVQ-LSIEANATVRSYALISGQPFVLCSTQVIGPSALETFCUNDEGRALLPQG 249
Qy      247 GGHRRITGPDGDLVAPLAENEGILYANLDPGVRIILAKMAADPAHYSRPDITRLIDR 306
       250 CGMARIVGPDGSELAKPLAEDAEGLYAEIDLEQILAKAGADPVGHYSRPDVLVSQFDP 309
Db      307 SPKLPV--VEIEGDL 319
       310 RNHTPVHRIIGIDRL 324

```

```

RESULT 9
US-11-096-568A-20687
; Sequence 20687, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nickolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 20687
; LENGTH: 333
; TYPE: PRT
; ORGANISM: Zea mays subsp. mays
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)-(333)
; OTHER INFORMATION: Ceres Seq. ID no. 12386686
US-11-096-568A-20687

```

```

Query Match      25.6%; Score 451; DB 7; Length 333;
Best Local Similarity 33.3%; Pred. No. 1.4e-35;
Matches 107; Conservative 68; Mismatches 118; Indels 28; Gaps 8;

Qy      6 KVACVOAAPIYMDLEATVDTKTIEMEEAARNNAFLIAPETWIPGYP-----WFLMLDSP 60
       13 RVTVOAASSVFYTPTPATLDKAEVLVAEAGYSGQLFPEVFGYVPHGSTFGLVVGNR 72
Db      61 AMAQPVROYHENSLELDGPOAKRISDAKRLGIMVTLGMSERVGTLYISQWFIGDND 120
       73 AKGKEDFOKYHAAISAIIVPGFEVSRLSALGKYCVFLVIGVEAAGYTLVNTVLSFDP 132
Qy      121 TIGARRRLKPTFVERTLFGEDSSSLAVFESVRLGGLCWEHLQPLTKYALYAQNEBI 180
       133 YLGNHRKQMPALIERFWGFGDSTTPYVDTPIGKAGALICENRRPILRTAMYGIEI 192
Db      181 HCAMPFSFLYNAAKALGPDVNVVAASRIYAVEGQCFVLASCAVSSQSMID-----MLCT 235
       193 YCA--FTVDCMPFW-----LSWTHIALGCGCFVLSACFCRKNYPPPEYTF 241
Qy      236 DDEK--HALLAGGHSRIIGPDGDLVAPLAENEGILYANLDPGVRIILAKMAADPA 292
       242 LEEPPSPESVYSCGG--SVIISPLGTVLAGPNYES--BALITADLDGEIVRAKFD 298
Db      293 HYSRPDITRLIDRSPKLPV 313
       299 HYSRPVLSLVKSDPK-PAV 318

```

```

RESULT 10
US-11-096-568A-20686
; Sequence 20686, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nickolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 20686
; LENGTH: 351
; TYPE: PRT
; ORGANISM: Zea mays subsp. mays
; FEATURE:
; NAME/KEY: misc feature
; LOCATION: (1)-(351)
; OTHER INFORMATION: Ceres Seq. ID no. 12386685

```



US-11-096-568A-20686

Query Match 25.6%; Score 451; DB 7; Length 351;

Best Local Similarity 33.5%; Pred. No. 1.5e-35; Matches 107; Conservative 68; Mismatches 118; Indels 28; Gaps 8;

QY 6 KVACVQAPRYMDLEATVDKTIEMBEAARNNAIIFPETWIPGV-----WFLMDSP 60  
 DB 31 RVTYVQASSVFYDTPALDRAEKI VALAAGSGLVFPFVGVGYPHSGTFLVGNRT 90  
 QY 61 AAMQVRYQYHENSLELDGPQAKRISDAKRLGIMVTLGMSERVGTLTYSQWETGND 120  
 DB 91 AKGKEDQKTHASAIIDVPGPEVSRSLAGKRYKFLVIGVERAGYTLTMTVLSFDDLK 150  
 QY 121 TIGARRLKPTFVTRTLFGSDGSSLAVERSVRGLGCMHELOPLTKALYAQNEI 180  
 DB 151 YLGHRRKVPMPALERVWGFQDGTIPVDPGKMAALCWEKRMPLRLTAMTAKGIEI 210  
 QY 181 HCAMPFSFLYPNAKALGPDVNAASRIYAVEGQCFVLASCALVSQSMID-----MLCT 235  
 DB 211 YCA--PTVDCMPTW-----LSSMTHALBEGCCTVLSHCQCRKKNTPPPETFCG 259  
 QY 236 DDEK--HALLAGGHSRIIGPDGDLVAPLAENEGILYANLDPGVRILAKMAADPAG 292  
 DB 260 LREBPSPEVYVCSG--SVIISPLGTVLAGPNYS--EALITADLDLGEIVRAKDFDVG 316  
 QY 293 HYSRPTRLIDRSPTLPV 313  
 DB 317 HYSRPEVLSLVKSDPK-PAV 336

RESULT 11

US-11-096-568A-6955

Sequence 6955, Application US/11096568A

Publication No. US20060048240A1

GENERAL INFORMATION:

APPLICANT: Alexandrov, Nikolai et al.

TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides

FILE REFERENCE: 2750-1592PUS2

CURRENT APPLICATION NUMBER: US/11/096, 568A

CURRENT FILING DATE: 2005-04-01

NUMBER OF SEQ ID NOS: 34471

SEQ ID NO 6955

LENGTH: 330

TYPE: PRT

ORGANISM: Glycine max

FEATURE:

NAME/KEY: misc feature

LOCATION: (1)-(330)

OTHER INFORMATION: Ceres Seq. ID no. 15169318

US-11-096-568A-6955

Query Match 25.3%; Score 446.5; DB 7; Length 330;

Best Local Similarity 33.5%; Pred. No. 3.8e-35;

Matches 111; Conservative 60; Mismatches 113; Indels 47; Gaps 10;

QY 5 IKVACVQAPRYMDLEATVDKTIEMBEAARNNAIIFPETWIPGYPMFLMDSPAM 64  
 DB 10 VRATVQASTIFDYDTPALDRAEKI VALAAGSGLVFPFVGVGYPHSGTFLVGNRT 90  
 QY 65 QF-----VRQYHENSLELDGPQAKRISDAKRLGIMVTLGMSERVGTLTYSQW 114  
 DB 65 SIGNRTVKGREERKTHSAIDVPGPEVSRSLAGKRYKFLVIGVERAGYTLTMTVLS 124  
 QY 115 IGDNGDTIGARRLKPTFVTRTLFGSDGSSLAVERSVRGLGCMHELOPLTKALY 174  
 DB 125 FDSQGHYLGHRKIMPTALERVWGFQDGTIPVDPGKMAALCWEKRMPLRLTAMT 184  
 QY 175 AQNEIHCAMPFSFLYPNAKALGPDVNAASRIYAVEGQCFVLAS--CA----- 223  
 DB 185 AKGVEIYCA-----PTAD--ARDVQASMTHTALBEGCCTVLSANQCRRDYPPPP 233

QY 224 --LVQSQMDMLCTDDEKHALLAGGHSRIIGPDGDLVAPLAENEGILYANLDPGVR 281

DB 234 EYFAGTEVDL--TPD--SVCAAG--SVIISPLGTVLAGPNYD--EALISADLDGEI 285

QY 282 ILAKMAADPAGHSRPTRLIDRSPTLPV 312

DB 286 ARAKFDVVGHSRPEVLSLVWDPHTPNV 316

RESULT 12

US-11-096-568A-6954

Sequence 6954, Application US/11096568A

Publication No. US20060048240A1

GENERAL INFORMATION:

APPLICANT: Alexandrov, Nikolai et al.

TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides

FILE REFERENCE: 2750-1592PUS2

CURRENT APPLICATION NUMBER: US/11/096, 568A

CURRENT FILING DATE: 2005-04-01

NUMBER OF SEQ ID NOS: 34471

SEQ ID NO 6954

LENGTH: 350

TYPE: PRT

ORGANISM: Glycine max

FEATURE:

NAME/KEY: misc feature

LOCATION: (1)-(350)

OTHER INFORMATION: Ceres Seq. ID no. 15169317

US-11-096-568A-6954

Query Match 25.3%; Score 446.5; DB 7; Length 350;

Best Local Similarity 33.5%; Pred. No. 4.1e-35;

Matches 111; Conservative 60; Mismatches 113; Indels 47; Gaps 10;

QY 5 IKVACVQAPRYMDLEATVDKTIEMBEAARNNAIIFPETWIPGYPMFLMDSPAM 64  
 DB 30 VRATVQASTIFDYDTPALDRAEKI VALAAGSGLVFPFVGVGYPHSGTFLVGNRT 90  
 QY 65 QF-----VRQYHENSLELDGPQAKRISDAKRLGIMVTLGMSERVGTLTYSQW 114  
 DB 85 SIGNRTVKGREERKTHSAIDVPGPEVSRSLAGKRYKFLVIGVERAGYTLTMTVLS 144  
 QY 115 IGDNGDTIGARRLKPTFVTRTLFGSDGSSLAVERSVRGLGCMHELOPLTKALY 174  
 DB 145 FDSQGHYLGHRKIMPTALERVWGFQDGTIPVDPGKMAALCWEKRMPLRLTAMT 204  
 QY 175 AQNEIHCAMPFSFLYPNAKALGPDVNAASRIYAVEGQCFVLAS--CA----- 223  
 DB 205 AKGVEIYCA-----PTAD--ARDVQASMTHTALBEGCCTVLSANQCRRDYPPPP 253  
 QY 224 --LVQSQMDMLCTDDEKHALLAGGHSRIIGPDGDLVAPLAENEGILYANLDPGVR 281  
 DB 254 EYFAGTEVDL--TPD--SVCAAG--SVIISPLGTVLAGPNYD--EALISADLDGEI 305  
 QY 282 ILAKMAADPAGHSRPTRLIDRSPTLPV 312  
 DB 306 ARAKFDVVGHSRPEVLSLVWDPHTPNV 336

RESULT 13

US-11-096-568A-6956

Sequence 6956, Application US/11096568A

Publication No. US20060048240A1

GENERAL INFORMATION:

APPLICANT: Alexandrov, Nikolai et al.

TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides

FILE REFERENCE: 2750-1592PUS2

CURRENT APPLICATION NUMBER: US/11/096, 568A

CURRENT FILING DATE: 2005-04-01

NUMBER OF SEQ ID NOS: 34471

SEQ ID NO 6956

```

; LENGTH: 233
; TYPE: PRT
; ORGANISM: Glycine max
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (1)..(233)
; OTHER INFORMATION: Ceres Seq. ID no. 15169319
US-11-096-568A-6956

```

```

Query Match      18.3%; Score 323; DB 7; Length 233;
Best Local Similarity 35.9%; Pred. No. 1.9e-23;
Matches 85; Conservative 39; Mismatches 81; Indels 32; Gaps 8;

```

```

QY 89 AKRLGIWITGMSRVSQGTIYISQWFIQDNGDTIGARKKLPFTFVERTLFGEGDSSILAV 148
DB 2 AGKYVHLVWGVIERDGTITVLFPSQGHYIGKRIKIMPTLRELVINGPQDGTITPV 61
QY 149 FETSVGLGGLCWEHLQPLTKVALYAQNEIHCAMPSESLYPNAKALGPDVNVAAASR 208
DB 62 FETVKGKIGALICWENMPLRTAMVAKGVEIYCA-----PTAD---ARDVWQASMT 110
QY 209 IYAVEGQCFVLAS---CA-----LVSQSMIDMLCTDDEKHALLAGGHSRIIGP 255
DB 111 HIALEGGCFVLSAQFCRRRDYPPPEYVFAGTEVDL--TPD--SVVCHGG--SVIISP 163
QY 256 DGGDLVAPLAENEGILYANIDPGVRILAKMAADPAGHSRPDITRLIDRSPLPV 312
DB 164 LGAVLAGPNYDG--EALISADLDLGEIARAKPDPVGHYSRPEVLSITVKDHPFNPV 219

```

```

RESULT 14
US-11-096-568A-20688
; Sequence 20688, Application US/11096568A
; Publication No. US20060048240A1
; GENERAL INFORMATION:
; APPLICANT: Alexandrov, Nickolai et al.
; TITLE OF INVENTION: Sequence-Determined DNA Fragments and Corresponding Polypeptides
; FILE REFERENCE: 2750-1592PUS2
; CURRENT APPLICATION NUMBER: US/11/096,568A
; CURRENT FILING DATE: 2005-04-01
; NUMBER OF SEQ ID NOS: 34471
; SEQ ID NO 20688
; LENGTH: 193
; TYPE: PRT
; ORGANISM: Zea mays subsp. mays
; NAME/KEY: misc_feature
; LOCATION: (1)..(193)
; OTHER INFORMATION: Ceres Seq. ID no. 12386687
US-11-096-568A-20688

```

```

Query Match      15.6%; Score 274.5; DB 7; Length 193;
Best Local Similarity 35.4%; Pred. No. 7e-19;
Matches 68; Conservative 36; Mismatches 65; Indels 23; Gaps 7;

```

```

QY 130 PTFVERTLFGEGDSSILAVFETSVGLGGLCWEHLQPLTKVALYAQNEIHCAMPSEFS 189
DB 2 PTHLEVFVWFGDGTITPVYDPIGKMGALICWENMPLRTAMVAKGIEIYCA--PTVD 59
QY 190 LYPNAKALGPDVNVAAASRIYAVEGQCFVLASGALVSQSMID-----MLCTDDEK---HA 241
DB 60 CMPTW-----LSMTHIALLEGCFVLASQFCRRKNYPPPEYTFCGLEEPSPEFS 110
QY 242 LILAGGHSRIIGPDDGDLVAPLAENEGILYANIDPGVRILAKMAADPAGHSRPDITR 301
DB 111 VVCSGG--SVIISPPLGTIVLAGPNYES--EALITRADLDLGEIARAKPDPVGHYSRPEVLS 167
QY 302 LILDRSPKLPVV 313
DB 168 LVVKSDEK--PAV 178

```

```

RESULT 15
US-11-079-463-5457
; Sequence 5457, Application US/11079463
; Publication No. US20060073161A1
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO BACTERIOIDES FRAC
; FILE REFERENCE: PATH00-03DIV2
; CURRENT APPLICATION NUMBER: US/11/079,463
; CURRENT FILING DATE: 2005-03-14
; PRIOR APPLICATION NUMBER: US 60/128,705
; PRIOR FILING DATE: 1999-04-09
; PRIOR APPLICATION NUMBER: US 09/540,209
; PRIOR FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 10444
; SEQ ID NO 5457
; LENGTH: 295
; TYPE: PRT
; ORGANISM: B. fragilis
US-11-079-463-5457

```

```

Query Match      7.1%; Score 125; DB 7; Length 295;
Best Local Similarity 22.7%; Pred. No. 0.00032;
Matches 78; Conservative 44; Mismatches 106; Indels 116; Gaps 19;

```

```

QY 5 IKYACVQAAPITMDLEATYDKTIELM-----EAAARNARLIAPETWIPGYPFWLMD 58
DB 5 IKVGIIQQA-----NTSDIRINIMLAKSIEACANGAHLVIOE-----LH 46
QY 59 SPAMAMQFVROYHENSELEL-----DGPQKRISD--AAKRLGIWITGMSRVSQGTIYISQ 112
DB 47 N---SYFCQENTDILFELAEPIPGSTGFYSBLAAARI--VLVTSLEPKRABGLYHNVA 102
QY 113 WFIGDNGDTIGARRKL---KPTFVERTLFGEGDSSILAVFETSVGLGGLCWEHLQPL 168
DB 103 VVPRDQSGIAGKYRKWHIPDDPAYEKFYFTPGD--IGPEPIQTSIGKGLVLCWQD--- 157
QY 169 TKVALYAQNEIHCAMPSESLYPNAKALGPDVNVAAASRIYAVEGQCFVLASGALVSQS 228
DB 158 -----W-----YPERA-----RLMALKAELIYPTLIGWES 184
QY 229 MIDMLCTDDEK---HALLAGGHSRIIG-----PDGDLVAPLAENEGIL 272
DB 185 -----TDTDEKRRQLNAMIISQRAHAVANGLPVYSVNRVGHEDP-----PSGQWGIILF 234
QY 273 YANIDPGVRILAKMAADPAGHSRPDITRLIDRSKLPVVEIE 316
DB 235 WGN-----SFVAGPQGEY-----LAQAGNDRSENMM--IVEVD 264

```

```

Search completed: April 27, 2006, 01:06:51
Job time : 18.7628 secs

```